Improving Preventable Disease Detection with Intelligent Remote Monitoring

Caroline Cocca

INTRODUCTION

In the United States, the privatized healthcare industry has created discrepancies between the treatment of privileged and unprivileged groups. Many Americans across the country lack access to consistent, affordable healthcare, both due to monetary and physical distance limitations. For example, in 2020, 13.9% of individuals aged 18-64 were uninsured [1], 11.4% of adults did not have a usual place to receive medical care [2], and 6.7% of adults failed to get needed medical care due to a lack of money [3]. Furthermore, because sick leave is not a guaranteed benefit in the United States [4], many working class citizens do not have the capability to miss work in order to make a doctor's appointment. In terms of distance, 18% of citizens live more than 10 miles from a hospital, and within the quarter of rural citizens surveyed by the Pew Research Center with the longest travel times, an average of 34 minutes was needed to get to an urgent medical care facility [5]. Clearly, there are pressing improvements needed both in the affordability of medical care, as well as of the capability to receive care that is restricted due to distance limitations.

Another issue in healthcare that plagues the nation is the prevalence of deaths due to preventable diseases that could have been circumvented with earlier intervention and proper care. In specific, according to the CDC, up to 40% of annual deaths from each of the five leading causes in the United States could have been prevented. These five leading causes include cander, stroke, heart disease, chronic lower respiratory diseases, and unintentional injuries [6].

The above discrepancies in the availability of healthcare have consequently created discrepancies in the effects of these preventable diseases on privileged and unprivileged populations.

DECOMPOSITION

Our current medical system makes needed care inaccessible for many citizens across the country, which is based both on financial and physical limitations. When patients can't receive care, this is a lose-lose situation for all: the patient's health suffers, the medical providers are missing out on patients they could have otherwise treated and been compensated for, and the government faces a nation with an incredibly large amount of healthcare spending that results from the snowballing process of preventable diseases not being caught early on. In order to address this problem, solutions must be investigated that can lower the cost of medical care as well as bridge patients with providers that are too far away to see regularly. A remote healthcare system that is enabled by AI to make patient monitoring less time-intensive for providers will both allow physically distant patients to be treated by these physicians as well as decrease the cost of this treatment through its improved efficiency.

This system thus results in a return of investment for the providers that will be able to see more patients in a more timely manner, as well as for the government that will benefit from decreased death due to preventable disease and decreased long-term healthcare spending. Furthermore, this system provides an invaluable worth to the patients who will gain access to potentially life-saving treatment and improved quality of life.

DOMAIN EXPERTISE

The AI-enabled portion of the system will entail predictive models that require training on medical data in order to identify patterns between a patient's vitals and medical history with a preventable disease. Consequently, input from medical professionals is crucial for every portion of the model development process. The preventable disease selection, gathering of appropriate and accurate data, and review of model explainability and accuracy all require input from medical professionals that can verify the trustworthiness and correct implementation of these steps in the pipeline. In specific, providers that are experienced with the exact preventable disease chosen will provide the most helpful review of the accuracy of the model, as well as the level of explainability needed to have their diagnoses *supplemented* by understandable model output, rather than made *for them* by a black-box system.

Furthermore, because this system is targeted towards underserved populations, members of these communities will provide crucial input on the design of the system in order to ensure user acceptance and trust. Engineering a remote healthcare system to aid a given population will be useless without the adoption of the system by patients within this population. This system must be user-friendly, explainable, and trustworthy in order for the population to seek medical care through it.

Finally, the domain expertise of healthcare systems engineers and security experts is needed to ensure the technical competency and safety of the system. Medical data is sensitive, personally identifiable information, which is protected under law [7] and requires safeguards in order to train a model on this data, as well as record and process ongoing collection of patient data. Having experienced input on this factor of the system is crucial to ensuring patient privacy. Input from healthcare systems engineers will also provide needed insight on proper implementation on predictive analytics of medical data, including data engineering, modeling, and storage.

DATA

The preventable disease chosen for the proof-of-concept model in the system project is diabetes. The Pima Indians Diabetes dataset, sourced from the National Institute of Diabetes and Digestive and Kidney Diseases [8], was selected as the training data for a proof-of-concept predictive model. The dataset contains 768 observations of a population of women (aged 21 and up) of Pima Indian heritage near Phoenix, Arizona. A combination of medical history and vitals are recorded, with the target variable being a binary indicator of the patient having diabetes. The features in this dataset are described in the table below, and include a variety of information from number of pregnancies to blood pressure and glucose levels. This dataset was selected due to diabetes being a preventable disease which was the eighth leading cause of death in the United States in 2020 [9], as well as due to the presence of vitals signs in the dataset that could be measured by remote monitoring devices, such as a blood pressure cuff, on regular intervals to be processed by the system's model in combination with the patient's previously recorded medical history.

Feature	Description
Pregnancies	Number of pregnancies experienced in the patient's lifetime
Glucose	Plasma glucose concentration obtained by an oral glucose tolerance test
Blood pressure	Diastolic blood pressure (mm Hg)
Skin thickness	Triceps skin fold thickness (mm)
Insulin	2-Hour serum insulin (mu U/mL)
BMI	Body mass index (weight in kg / (height in m)^2)
Diabetes Pedigree Function	A likelihood score of diabetes based on family history
Age	Age in years

Outcome	Indication of diabetic (1) or not diabetic (0)

Figure: Description of the features in the Pima Indians Diabetes dataset.

Because this dataset is limited to a small population of citizens, the consequently trained model may not generalize well to predicting diabetes in other populations due to a domain shift. However, this model is meant to serve as a proof-of-concept demonstration of a preventable disease model, and in public deployment, the training data would need to be expanded depending on the populations receiving this remote care. A model trained on a limited dataset such as this one would also be suitable for a limited pilot test within that specific population.

In order to ensure data quality, the dataset was scanned for NaN values. The dataset, unfortunately, has an unusual codification of missing values by using 0s instead of NaNs. These values were present in the glucose, blood pressure, insulin, skin thickness, and BMI columns, which were consequently imputed with the mean values of that feature. There were no missing values from the diabetes pedigree and age columns, though missing values in the pregnancy column were unclear due to 0 being a valid value. It's of note that the insulin column had a very high number of null values (374, or 48.6% of observations), as did skin thickness (227, or 29.6% of observations). Imputing this amount of values may result in an altered distribution of the feature that is not representative of its true distribution in the population, and thus could skew classification results.

Furthermore, all columns were normalized using the MinMaxScaler transform of the sklearn library to ensure all features were on the same scale. It was not necessary to convert any columns with tools such as one-hot encoding due to all features being numeric.

Of final note, there is an underrepresentation of positively classified observations (268, or 34.9% of observations) compared to negatively classified observations (500, or 65.1% of observations.) This class imbalance will require close attention to metrics other than accuracy, such as F1 score and confusion matrices, in order to ensure that an accuracy score is not trivially obtained by model bias towards a negative outcome.

DESIGN

Feature exploration was performed in order to gain an initial understanding of correlation between features and the outcome variable, as well as to potentially perform feature elimination. This exploration was done via observing histograms of the features, a correlation matrix (shown in the figure below), as well as training a random forest classifier on the dataset in order to access its feature importance attribute to observe an initial ranking of the features. Both the correlation matrix and random forest feature ranking indicated that glucose was the strongest predictor of diabetes, followed by BMI and age. However, there were some differences between the correlation and RF ranking, such as the diabetes pedigree function showing the lowest correlation with the outcome variable (tied with blood pressure), but received the fourth highest ranking by the RF classifier. Because the list of features is already quite small compared to many other datasets, and none were consistently ranked at the bottom in either examination, no features were excluded from the model selection process.

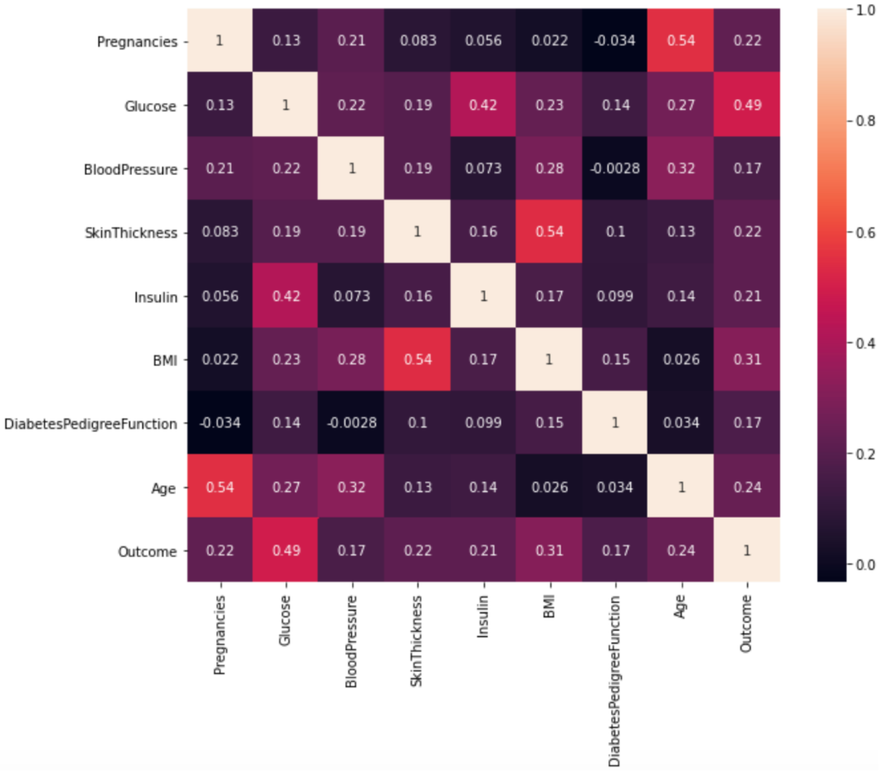


Figure: A correlation matrix of the Pima Indians Diabetes dataset.

Due to the importance of model explainability in order for the outputs to appropriately augment physician diagnoses, as well as of accuracy to ensure trustworthy flagging of potential diabetic patients, a model selection process was conducted to determine the best balance between explainability and performance on the dataset. The algorithms evaluated in the model selection process include Decision Tree, Random Forest, k-Nearest Neighbors, and Naive Bayes. For each model, a hyperparameter was also chosen in order to tune the value by training multiple instances of the model with varying values for that parameter. The parameters include the max depth of the decision tree, the number of estimators in random forest, the number of neighbors in k-NN, and the smoothing hyperparameter of naive bayes.

DIAGNOSIS

The models were evaluated through 5-fold cross validation, and the results of this evaluation are displayed in the table below, which include the hyperparameter selected, accuracy, and macro F1 score of each algorithm. The final selected model was ultimately k-NN with the number of neighbors hyperparameter tuned to 30. Although Random Forest with n_estimators set to 150 had a slightly better macro F1 score (0.727) compared to 30-NN (0.726), the 30-NN model had an accuracy (0.766) that was significantly higher than the RF model (0.759), and so the accuracy was prioritized.

Model	Chosen hyperparameter	Accuracy	Macro F1
k-NN	n_neighbors = 30	0.766 ± 0.031	0.726 ± 0.035
Naive Bayes	var_smoothing = 0.01	0.743 ± 0.029	0.713 ± 0.028

Decision Tree	max_depth = 2	0.732 ± 0.038	0.697 ± 0.040
Random Forest	n_estimators = 150	0.759 ± 0.034	0.727 ± 0.039

Figure: Model evaluations for k-NN, Naive Bayes, Decision Tree, and Random Forest algorithms and their selected parameters.

Although the accuracy and F1 scores are significantly better than the 50% accuracy a randomly guessing model would have, these scores are still low. There are several factors influencing the difficulty of model training and selection for this problem. Firstly, this is a relatively small dataset containing only 768 observations. There may simply not be enough training data in this set in order for the model to have enough information to learn an accurate, generalizable set of parameters. Furthermore, the dataset also has a limited number of features. Although some of these features are ranked significantly high in usefulness as predictors of the target variable (such as glucose and BMI), there are likely features missing that also contain valuable information on the outcome variable that the model does not have the chance to learn. Even within the features included in the dataset, some had a large amount of missing values, such as insulin, which required value imputation that could potentially create a feature distribution that is not accurate to the true distribution in the population.

As with many medical datasets, the number of positive examples of the outcome variable are significantly less than that of negative examples. This further contributes to a lack of sufficient information for the model to successfully train, as well as presents a bias risk towards negative examples due to their overrepresentation. The macro F1 score was monitored for this reason, which uses an unweighted mean of each class, and a confusion matrix was also observed. These evaluation metrics showed that the model was not gaming the accuracy score through excessive classification of the negative class. However, as with the accuracy, this score could be further improved by collecting more samples from the positive class.

Finally, medical diagnosis problems are often difficult in and of themselves for models to learn [10]. This space is not as thoroughly explored and iterated upon as other subsets of AI, such as computer vision. The complexity of these diagnoses is the precise reason why explainable models must be prioritized in medical applications so that a trained human can double check these results and augment them with their own experience.

DEPLOYMENT

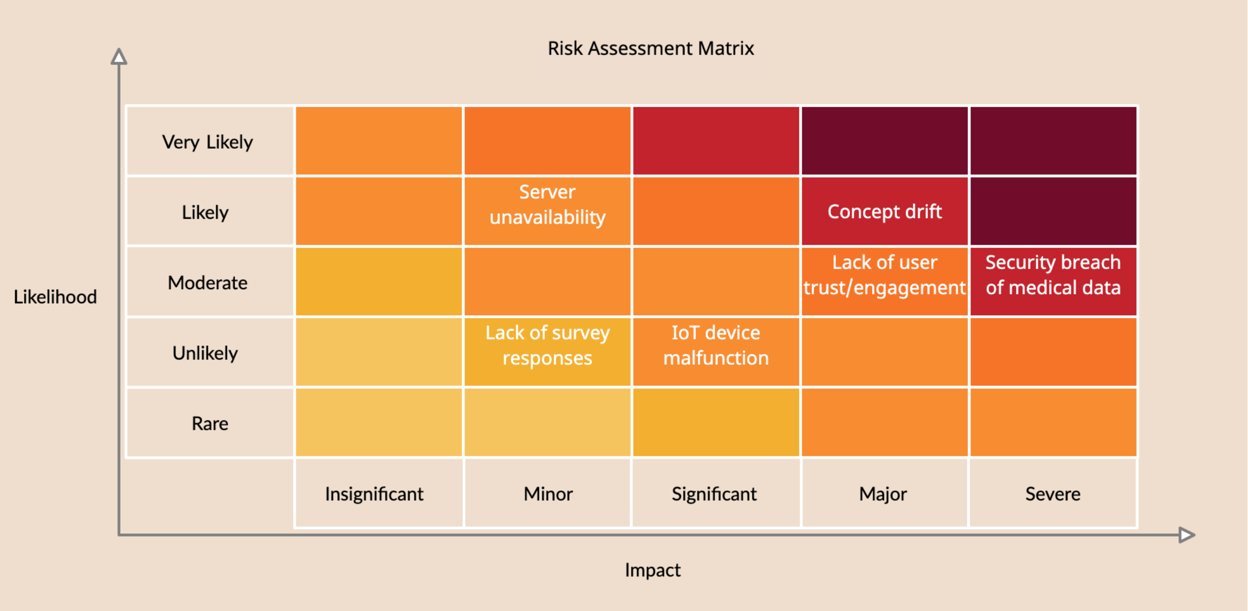
A system utilizing the above algorithm for remote monitoring of diabetes would require a pipeline in order to intake and process medical readings, feed these into a model, and output the flags for a medical provider to review. This pipeline necessitates several key components: IoT devices for medical measurements, bluetooth or wired transmission of this data, a relational database for storage of this data, and a messaging component to feed readings into the model and save results back to the database.

On the server-side, data would be received and stored from IoT devices, which would then be used to iteratively fine-tune the model. Because these measurements are not being viewed in real-time by a physician and will only be making updates perhaps daily or weekly, these measurements would also be processed server-side by the model and stored back into the database as a flag column value in that observation’s row. On the client-side, a provider could then query data by patient or date in order to review recent flags or check up on specific patients as needed.

Continuous monitoring would be required both for model performance validation as well as for security. If the model is experiencing concept drift due to a change in distribution in the population, or is perhaps overfitting through its iterative training, data scientists must be able to oversee and correct these issues through model versioning and iterative development. Furthermore, the system must have trustworthy, continuous security monitoring in order to ensure the safety of HIPAA-protected medical data. Because a breach has potential significant consequences for patient privacy, as well as for violation of the law, security must be a heavily prioritized component of the system. This would include logging of all access to the system, multi-factor authentication systems, and defense against client-side exploitation of the interface that could result in a data spill.

Lastly, following up on deployment through surveys given to both patients and providers will be crucial to evaluating how intuitive, trustworthy, and useful the system is for both parties. This survey data could be used to collect feedback on potential improvements, such as changes requested to the user interface by providers, as well as flag causes for concern with user engagement before a significant loss of engagement occurs.

A risk matrix is shown below in order to illustrate the above risks, as well as note some further minor risks associated with deployment.



	Insignificant	Minor	Significant	Major	Severe
Very Likely					
Likely		Server unavailability		Concept drift	
Moderate				Lack of user trust/engagement	Security breach of medical data
Unlikely		Lack of survey responses	IoT device malfunction		
Rare					

Figure: A risk assessment matrix for system deployment.

CONCLUSION

Through the usage of intelligent, predictive algorithms, remote healthcare can be made more efficient for providers, and consequently more accessible for their patients. However, further refinement would be required for model deployment in order to ensure that providers are using the most accurate, highest quality model output possible in order to augment their diagnoses. A k-Nearest Neighbors model trained on the Pima Indians Diabetes dataset showed promise in the possibility of the system being able to integrate a fully explainable, intuitive algorithm that could be easily understood by doctors with no formal education in machine learning. Moving forward, all predictive model development for the system must continue to prioritize explainability in order to avoid black-box decision output that could cause serious harm to patients without the medical provider’s ability to understand and analyze this output. Furthermore, there are risks outlined above that require further mitigation plans before deployment can occur, such as security design, gathering of domain expert knowledge on how to ensure user engagement for specific populations, and planning an iterative development process on predictive modeling in order to prevent concept drift and continuously improve model performance. Although predictive modeling of medical conditions is a difficult problem, it is a worthy one of further exploration in order to provide healthcare to underserved populations and work towards decreasing deaths due to preventable disease.

References

[1] National Center for Health Statistics, “Health insurance coverage: Early release of estimates from the National Health Interview Survey, 2020,” CDC, 2021, https://www.cdc.gov/nchs/data/nhis/earlyrelease/insur202108-508.pdf

[2] National Center for Health Statistics, “Percentage of having a usual place of health care for adults aged 18 and over,” United States, 2019—2020. National Health Interview Survey. Generated interactively: Apr 24 2022 from https://wwwn.cdc.gov/NHISDataQueryTool/SHS_adult/index.html

[3] National Center for Health Statistics, “Early release of selected estimates based on data from the 2020 National Health Interview Survey,” CDC, 2021, https://www.cdc.gov/nchs/data/nhis/earlyrelease/EarlyRelease202108-508.pdf

[4] U.S. Department of Labor, “Sick leave,” 2022, https://www.dol.gov/general/topic/workhours/sickleave

[5] O. Lam, B. Broderick, and S. Toor, “How far Americans live from the closest hospital differs by community type,” Pew Research Center, 2018, https://www.pewresearch.org/fact-tank/2018/12/12/how-far-americans-live-from-the-closest-hospital-differs-by-community-type/

[6] U.S. Department of Health and Human Services, “Up to 40 percent of annual deaths from each of five leading US causes are preventable,” CDC, 2014, https://www.cdc.gov/media/releases/2014/p0501-preventable-deaths.html

[7] Center for State, Local, Tribal, and Territorial Support, “Health Insurance Portability and Accountability Act of 1996 (HIPAA),” CDC, 2018, https://www.cdc.gov/phlp/publications/topic/hipaa.html

[8] J. Smith., J. Everhart., W. Dickson, W. Knowler, and R. Johannes, “Using the ADAP learning algorithm to forecast the onset of diabetes mellitus,” Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261--265), IEEE Computer Society Press, 1988.

[9] National Center for Health Statistics, “Deaths and Mortality,” CDC, 2020, https://www.cdc.gov/nchs/fastats/deaths.htm

[10] P. Desikan, R. Khare, J. Srivastava, R. Kaplan, J. Ghosh, L. Liu, V. Gopal, “Predictive Modeling in Healthcare: Challenges and Opportunities,” IEEE Life Sciences Newsletter, 2013, https://lifesciences.ieee.org/lifesciences-newsletter/2013/november-2013/predictive-modeling-in-healthcare-challenges-and-opportunities/