Medical Decision Support System: Diabetes Diagnosis and Treatment Recommendations

Computer Science, CSU Global

Foundations of Artificial Intelligence

Dr. Pubali Banerjee

M. Danish

June 11, 2023

For this project, I designed a Medical Decision Support System made up of a classifier component and a decision support component. My Medical Decision Support System analyzes data from diabetic and non-diabetic patients and uses that analysis to predict whether a new patient is likely to be diabetic. Although I am not a medical professional, I have a small amount of domain knowledge on the subject of diabetes type 2 due to my own family medical history and personal experience with being tested and then treated for pre-diabetes as a co-occurring condition of Polycystic Ovarian Syndrome. The dataset I used in this project was interesting to me on a personal level and contained features with which I already have some familiarity.

Tools, Libraries, Deep Learning Methods, and Symbolic Planning

The Diabetes Mellitus Decision Support System uses the Scikit Learn, NumPy, and pandas libraries. Originally, I had planned to use PyLog to construct my Decision Support function. Ultimately, however, I decided against using PyLog due to the nature of the diabetes dataset I used. Because the diabetes dataset only produces a binary outcome—i.e., the patient either does have diabetes (1) or does not have diabetes(0)—and PyLog is designed to work with nuanced decision making logic (Huang et al., 2021), I decided that it would not be necessary to use PyLog. Instead, I wrote my own Decision Support System function that provides the physician with one of two sets of further information about the patient's condition. Due to the binary outcome value provided by the dataset and decision tree classification system, it also was not necessary for this program to invoke symbolic planning (Speck et al., 2019).

For its patient diagnosis recommendation, the DM-DSS uses the Scikit Learn Decision Tree Classification system (*Sklearn.tree.DecisionTreeClassifier*, 2023). First, the physician is presumed to have updated the existing dataset in the program with the patient's lab results and other information (Table 1). Next, the DTC model splits the total dataset into training and testing subsets. It uses the training set to train the model, then uses the testing subset to predict whether the patient has diabetes. The patient's data is always in the last row of the dataset, so the DSS function takes the last outcome value of `y_pred` and uses it to decide which set of recommendations to provide to the patient's physician. Because patient diagnoses would be handled separately due to HIPAA (AAPC, 2020; Bowie et. al, 2014), and there is only one set of patient results for the DM-DSS to report on at a time—presumably, in a laboratory setting—it was not necessary to use a search function in this version.

Representation of Knowledge and Expert Systems Concepts

In the previous section, I explained that the diabetes dataset only produces a binary outcome. The table below contains a sample of the diabetes dataset to further show how it represents patient data.

Table 1

Diabetes Dataset

Number of Pregnancies	Glucose (mmol)	Systolic Blood Pressure	Skin Thick-ness (mm)	Blood Insulin (mmol)	BMI	Diabetes Pedigree Function	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	21	0
3	154	80	30	94	32.5	0.400	32	0

Note: The last row represents the patient data I fabricated to test the program. The first three rows are taken from the Diabetes.csv dataset file.

In Table 1, the Number of Pregnancies column is self-explanatory—it represents the number of pregnancies the patient has had. Glucose represents their most recent blood glucose reading in mmol (National Institute of Diabetes and Digestive and Kidney Diseases, 2023). Systolic blood pressure represents the top number from a blood pressure reading (Pulse Pressure: An Indicator of Heart Health?, 2022). Skin thickness is a measurement that can indicate the progression of diabetes, especially type 1 (CDC, 2022). Blood insulin is the level of insulin in mmol that was detected in the patient's blood—i.e., a lab result (National Institute of Diabetes and Digestive and Kidney Diseases, 2023). BMI represents the patient's body mass index as calculated from their height and weight. The Diabetes Pedigree Function is a calculation that represents a patient's family history of diabetes (Das et al., 2022). Age is the patient's age in years at their last birthday. The outcome column represents the patient's diabetes diagnosis status, and as stated in the previous section, only a value of 1 represents a diabetes diagnosis.

In addition to the patient data and diagnosis information contained in the diabetes dataset, the DM-DSS program contains knowledge about two treatment paths. If the patient's outcome column value is 1, then the DM-DSS prints out a recommended initial treatment protocol for patients with Type 2 Diabetes Mellitus that I based on one authored by the Clinical Director at Massachusetts General Hospital's Diabetes Center—and Harvard Medical School Associate Professor of Medicine—Dr. Deborah J. Wexler (2022). If the patient's outcome column value is 0, then the DM-DSS advises the physician that the patient probably does not have diabetes at the time of its analysis, but that the physician should continue to follow-up with the patient regularly if their family history suggests a high likelihood of developing diabetes.

Conclusion

I think that with further development, a system like this could be extremely beneficial to diabetic patients. The recommended treatment protocol in the program advises physicians not to focus solely on A1C as a diabetes diagnosis metric (Bonora & Jaakko Tuomilehto, 2011), and not to ignore the potential for related neurological and cardiac issues that can develop in patients whose A1C might otherwise appear normal. This is a real risk diabetic patients face when treated by physicians who lack sufficient familiarity with diabetes to be able to recognize the potential for these problems. Catching these issues early via impartial statistical analysis, machine learning, and expert systems guidance could significantly improve quality of life for patients who might otherwise be at a disadvantage due to provider bias (e.g., racism or sexism), lack of diabetes expertise, and other factors that can negatively impact their care (Sutton et al., 2020).

References

AAPC. (2020). *HIPAA Reference Guide - First Edition*. AAPC. https://play.google.com/store/books/details?id=hNfbDwAAQBAJ

Bonora, E., & Jaakko Tuomilehto. (2011). *The Pros and Cons of Diagnosing Diabetes With A1C*. *34*(Supplement_2), S184–S190. https://doi.org/10.2337/dc11-s216

Bowie P, Halley L, McKay J. (2014, January 14). Laboratory test ordering and results management systems: a qualitative study of safety risks identified by administrators in general practice. *BMJ Open* 2014;4:e004245. doi: 10.1136/bmjopen-2013-004245

CDC. (2022, June 14). *Diabetes and Your Skin*. Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/library/features/diabetes-and-your-skin.html#:~:text=If%20blood%20sugar%20levels%20remain,have%20high%20blood%20sugar%20levels.

Das, M., Bhattacharyya, G., Gong, R., Misra, R., Medda, S. K., Banik, S., & Das, R. N. (2022). Determinants of Gestational Diabetes Pedigree Function for Pima Indian Females. *Internal Medicine – Open Journal*, *6*(1), 9–13. https://doi.org/10.17140/imoj-6-121

Huang, S., Wu, K., Jeong, H., Wang, C., Chen, D., & Hwu, W. W. (2021). *PyLog: An Algorithm-Centric Python-Based FPGA Programming and Synthesis Flow*. 1–1. https://doi.org/10.1109/tc.2021.3123465

Pedregosa et al. (2011). Scikit-learn: machine learning in Python. *JMLR 12*, pp. 2825-2830.

Pulse pressure: An indicator of heart health? (2022). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/expert-answers/puls e-pressure/faq-20058189#:~:text=Blood%20pressure%20readings%20are%20given,between%20beats%20(diastolic%20pressure).

sklearn.tree.DecisionTreeClassifier. (2023). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

Speck, D., Geißer, F., & Torralba, R. M. A. (2019). *Symbolic Planning with Axioms*.

National Institute of Diabetes and Digestive and Kidney Diseases. (2023, June 12). *Insulin Resistance & Prediabetes*. National Institute of Diabetes and Digestive and Kidney Diseases; NIDDK - National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). *An overview of clinical decision support systems: benefits, risks, and strategies for success*. *3*(1). https://doi.org/10.1038/s41746-020-0221-y

Wexler MD MSc, D. (2022, October 31). *Initial management of hyperglycemia in adults with type 2 diabetes mellitus*. Uptodate.com. https://www.uptodate.com/contents/initial-management-of-hyperglycemia-in-adults-with-type-2-diabetes-mellitus/print