



IJCRR

Section: Healthcare

ISI Impact Factor
(2019-20): 1.628

IC Value (2019): 90.81

SJIF (2020) = 7.893



Copyright@IJCRR

Machine Learning Implementation and Challenges: A Study of Lifestyle Behaviors Pattern and Hba1c Status

Christie Natasha Archie¹, Debashish Das^{2,*} Fatemeh Meskaran³

¹School of Computing & Technology, Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia; ²School of ICT, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, Selangor Darul Ehsan, Malaysia.

ABSTRACT

Background: Diabetes is a chronic metabolic disease that has a long-term impact on the individual's well-being and one of the causes of adulthood death.

Objective: This research paper represents an attempt to find the correlation between lifestyle behaviour patterns and diabetes by leveraging machine learning in the form to facilitate patients with risk stratification in a population.

Results: The major findings that emerged were as follows: an unhealthy lifestyle and dietary pattern lead to Non-communicable Disease (NCD) including diabetes. In the form to identify diabetes, Glycated Hemoglobin (HbA1c) will be used to diagnose diabetes considering its efficiency and convenience to the patient. Furthermore, contrary to what has been assumed of the superiority of machine learning has been provided in many aspects, few challenges should be taken into consideration when it comes to the implementation of Machine Learning in the healthcare field, racial bias, for instance.

Conclusion: In the Asia Pacific region, there is a range of cut-off point of HbA1c values due to HbA1c is subject to external factors such as race and ethnicity. Therefore, narrowing down the population scope in healthcare is considered in this paper as the best practice to facilitate better accuracy and assurance.

Key Words: Diabetes, Glycated Hemoglobin, HbA1c, Lifestyle Behavior, Machine Learning, Noncommunicable diseases

INTRODUCTION

The global efficiency improvement in the human living condition gives an abundance transformation to lifestyle behaviour and food consumption pattern which might lead to various states of illness including cancer, cardiovascular, respiratory disease, hepatic illness, and diabetes.¹ According to the International Diabetes Federation (IDF), one of the causes of adulthood death is diabetes.²

Diabetes is a chronic metabolic disease that has a long-term impact on the individual's well-being.³ In 2019, IDF has reported that 463 million people are diagnosed with diabetes in 2019, which is 9.3% of the world population and it has been predicted that diabetes might increase to 578 million by 2030.² Additionally, several studies have been proven that unhealthy lifestyle behaviours during early adolescence can lead to obesity due to unhealthy diet, lack of physical exercise, high blood pressure which are the early symptoms of chronic diseases including diabetes.^{1,3}

American Diabetes Association has revealed that unlike Fasting

Plasma Glucose (FPG) and Oral Glucose Tolerance Test (OGTT) diagnostic method, Glycated Hemoglobin or also known as HbA1c is generally used as a measurement for diabetes that provides convenience where the patient does not require to fast, unaffected with daily stress and disease disruption and better preanalytical stability.⁴ HbA1c test is a standardized accurate measurement with international reference principles where 6.5% of HbA1c is suggested as the cut-off point for diabetes diagnostics.^{5,6}

Additionally, it is true that over the past decade, almost every aspect of modern life is somehow changed by the advancement of the machine to make predictions⁷. By leveraging Machine learning in healthcare, this research is focusing on identifying the correlation between lifestyle behaviour and dietary pattern and probability of diabetes by using Glycated Hemoglobin (HbA1c) as a diabetes measurement.

Corresponding Author:

Debashish Das, School of ICT, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, Selangor Darul Ehsan, Malaysia.

Email: debashish.das@xmu.edu.my

ISSN: 2231-2196 (Print)

ISSN: 0975-5241 (Online)

Received: 20.10.2020

Revised: 28.12.2020

Accepted: 18.02.2021

Published: 22.06.2021

THE LIFESTYLE BEHAVIOR PATTERN

Noncommunicable diseases (NCDs) including respiratory disease, chronic cancer, cardiovascular, and diabetes can be prevented by living in healthy lifestyle behaviour patterns.⁸ There are four lifestyle behaviour factors such as lack of dietary ingredients intake, physical inactivity, smoking habit, and alcohol consumption can lead to a range of health issues such as obesity and high blood pressure at the early age, which can give the rise of NCDs possibility in adulthood.⁹

Based on these premises, minimizing the risk for Noncommunicable diseases especially at an early age increase the opportunity for a better individual's well-being. A lifestyle and food consumption study on university students has been conducted by El-Kassas and Ziade revealed that overall population dietary consumption behaviour falls below the average of recommended levels and it is estimated that 41.2% among 369 students having low physical activity level due to significant living condition changes in the university environment. Regression analysis by this study showed that food cravings, increased appetite, comfort eating, parental obesity, and stressful eating have a high level of relationship with the probability of NCD disease including diabetes. Over the past decade, westernized food consumption which rich in fat, salt, sugar, and alcohol is becoming the substitute for healthy food that consists of dietary ingredients including fruit, vegetables, and fibre.¹

Furthermore, an investigation has been conducted by Al-Nakeeb et al. that there are few main factors of dietary options mostly derived from the individual itself. Dietary knowledge, food preferences and personal state of mind for instance. These become the most factor of how individual dietary practice. His studies believe that lifestyle and nutritional habits planning strategy in early adulthood should be taken into action in the form of lower the risk of a range of chronic illnesses. For instance, Obesity Intervention in Teenagers leveraging several disciplinary Intervention Mapping policies to introduce health awareness program for young adult¹⁰ considering that obesity is one of the common symptoms that lead to diabetes.¹¹ This concept can be applied by universities to increase the awareness of a healthy lifestyle and dietary programs for university students.¹²

DIABETES MEASUREMENT

Numerous health associations around the world have suggested some diabetes diagnosis tests which are: (1) Oral Glucose Tolerance Test (OGTT), (2) Fasting Plasma Glucose (FPG), and (3) Glycated Hemoglobin.^{13,14} FPG is used to measure a patient blood sugar level which requires at least 8 hours of fasting. While OGTT is a glucose medical test

that requires blood samples to assess the capability of blood to extract glucose in the body. A test of diabetes is done on a two-hour test with 75-gram oral tolerance glucose (OGTT) to be taken by the patient.¹⁵ Lastly, Glycated haemoglobin is also known as Hemoglobin A1c (HbA1c) which is a body glycosylated protein that represents an average blood glucose level which has 100 to 120 erythrocyte (two or three months) lifespan.¹⁶

Diagnosis cut off point for diabetes diagnostic criteria recommended by the American Diabetes Association¹⁵ and the World Health Organization⁶ is listed in **Table 1**:

Table 1: Three diabetes diagnostic criteria recommended by the American Diabetes Association [15] and the World Health Organization [6].

<i>Parameter</i>	<i>Cut off Value of Diabetes Measurement</i>
Fasting Plasma Glucose (FPG)	≥ 7.0 mmol/L (126 mg/dl)
Oral Glucose Tolerance test (OGTT)	≥ 11.1 mmol/L (200 mg/dl)
Glycated Hemoglobin (HbA1c)	≥ 6.5% (48 mmol/mol)

HbA1c Level for Diabetes Measurement

Although there are some diabetes measurement has been suggested by international references, HbA1c level can be the best alternative to assess the potentials of diabetes.¹⁶ Not only that, it is used for tracking glucose level for long-term treatment for the patient who has been diagnosed with diabetes as well.¹⁶ The benefit of using HbA1c for diabetes diagnosis has been highlighted by Sacks and the American Diabetes Association which is the capability of the test to be completed at any time of the day and no fasting is required to be done. Although HbA1c introduces the robust stability result, this measurement has not been worldwide adopted.¹⁷ Because HbA1c is subject to well-known factors outside the control of healthcare providers and should be taken into consideration to prevent arbitrary comparisons such as race, ethnicity, age, diabetes length and type, patient adherence, and comorbidity.¹⁸

As the result, HbA1c is adopted differently, in the Asia-Pacific region, Region like Malaysia and Singapore which is multi-racial country are advised to concern with racial differences Malaysia use HbA1c with a cut-off of 6.30%, while Singapore uses a cut-off between 6.10%-6.20%. On the other hand, New Zealand uses a cut-off of 6.70%, Thailand does not use HbA1c to diagnose diabetes. Finally, in Japan, a 6.00% cut-off was suggested for the Japanese population to be used for first stage screening with the sensitivity of 0.837 to 0.876 with other glu-

cose-based validation like Fasting Plasma Glucose (FPG) and oral glucose tolerance test (OGTT).¹⁷

Due to the efficiency and convenience that HbA1c provided as a diabetes measurement, HbA1c will be used as the target variable to predict patient who has potentially diagnose with diabetes. With that being said, appropriate sample data selection in a population has to be narrow down into racial scale which in this paper will be the focus on the Japanese population.

MACHINE LEARNING IMPLEMENTATION IN HEALTHCARE

With adequate data, machine learning leveraging mathematical concepts and statistical models to learn the pattern of the given dataset without explicit rules or instruction to perform prediction or classification tasks.¹⁹ The fundamental transformation of conventional biostatistics leads to the implementation of machine learning in clinical analysis.²⁰ Machine learning implementation requires huge amount of clinical data to be carefully labelled and organized. Medical data can be derived from clinical images, log data from medical devices, laboratory reports, or the form of doctor's notes.²¹

Guidance on a medical device issued by Food and Drug Administration (FDA) also suggested that the AI and Machine Learning instruments are required for clinical analytical, health validation and risk stratification.²² One of the machine learning techniques such as incremental learning has been widely used in healthcare as its capability to be iteratively enhanced the prediction accuracy.²³ This process leveraging the data feedback loop to improve the accuracy of the prediction of the system through continuously training as shown in **figure 1**.

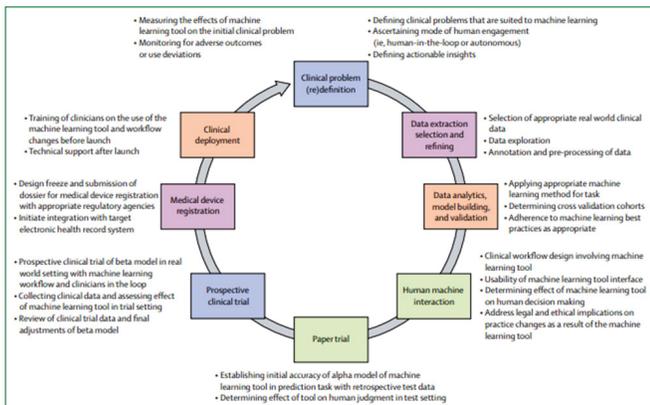


Figure 1: Healthcare implementation of Machine Learning using Incremental Learning.

There is a range of incremental learning models which are often used in the healthcare sector to make a prediction

which is: (a) Random Tree Forest (b) Multi-Layer Perceptron (MPL) (c) Support Vector Machine (SVM) (d) Neural Network that supports incremental learning.²³

Char et al. revealed that leveraging machine learning in complicated clinical practices is a constant issue, since the correct diagnosis and best practice may be problematic in a given case. The study believes that early incorporation in an algorithm of a specific diagnosis or procedure can indicate a legitimacy not supported by data.²⁴ The delivery of machine learning in the medical can vary by age, ethnicity, or race. Some of the problems might not draw a clear boundary of external non-genetic factor, and it is not protected from such concerns that algorithm may reflect human biases in making decisions. Several non-genetic studies in certain populations resulted in an algorithm designed to predict biased prediction²⁵. For instance, attempts to use the Framingham Heart Study dataset for cardiovascular risk prediction in non-white populations have provided limited results, both with overestimation of risk and underestimation of risk.²⁴ Therefore, it is important to narrowing the scope of prediction to prevent any external factor bias during model development.

Human to Machine Decision Making in Healthcare

Unlike Finance and Sales, the placement of the machine spectrum not always indicate superiority since healthcare cases and problem background require a different level of human guidance and involvement. Furthermore, every medical data has various complexity and variety. Contrary to what has been assumed of the superiority of machine learning has been provided in many aspects, few challenges should be taken into consideration when it comes to implementing Machine Learning in the healthcare.

Healthcare algorithm refinement requires continuous training to gain accurate outcome, not only that human involvement which is in the fourth phase of incremental learning as shown in **Figure 1** play a huge role to develop a reliable prediction²¹. Furthermore, the different algorithm used in healthcare has a different level of human involvement and machine to decide as shown in **figure 2**.⁷ The algorithm selection of each healthcare case is varying depending on the problem that needs to be addressed. For instance, Atrial Fibrillation stroke prediction uses a regression algorithm.^{25,26} that requires a high level of human effort to make decision making. While Random forest is used to predict Cardiovascular risk²⁷ which that require moderate human guidance. Finally, the application of Neural Network requires less human decision making to predict diabetic retinopathy using the image recognition concept.²⁸

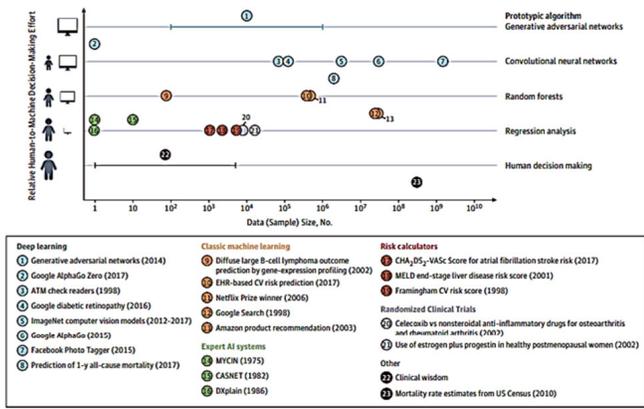


Figure 2: Human and Machine Decision Making²⁸.

RELATED WORKS ON HBA1C STATUS PREDICTION

Recent work by Karpati et al. proposed K-mean clustering to identify the clusters of patients and facilitate the characteristic of each patient cluster for better treatment. The outcome resulted in three clusters of HbA1c level which are: (a) stable, (b) low, (c) high with 99.8 % accuracy.^{28,29} On the other hand, the work by Parigi et al. proposed an analysis of HbA1c statue to targeting Haemoglobin A1c below 7% for diabetes patients. Random Forest model is used for model development to analyze the diabetes treatment response and patient lifestyle.³⁰ Finally, Nagaraj et al. proposed three supervised learning models which are: (a) Support Vector Machine (b) Random Forest (c) Generalized Linear to develop predictive model in the form to determine Haemoglobin A1c response after insulin treatment and make comparison to determine the best machine learning approach. HbA1c measurement test result is used as the benchmark. The outcome of their work resulted that Linear regression has better performance compared to other models.³¹

CONCLUSION

Through domain research, previous researchers revealed that there is a correlation between lifestyle and dietary behavior with Noncommunicable Disease (NCD), including diabetes¹². Machine learning implementation can be used to predict HbA1c status from an individual’s lifestyle behaviours pattern, it will help to gain an awareness of living a healthy lifestyle. A small change in lifestyle habits, especially smoking habit, involved in physical activities, and consume dietary fibers, could reduce the prevalence of diabetes.

However, the appropriate application of Machine Learning in the healthcare sector has to be carefully planned considering that clinical cases require different levels of human guidance and involvement to make medical decisions. Algorithm

selection is also a crucial task, as each healthcare case is varying depending on the problem that needs to be addressed.

As the adoption of HbA1c to diagnose diabetes has not widely used especially in the multi-racial population in the Asia Pacific region due to the various range of cut-off values of HbA1c. Glycated Haemoglobin (HbA1c) can still be an alternative in the form to diagnose diabetes considering its convenience and effectiveness compare to Fasting Plasma Glucose (FPG) and oral glucose tolerance test (OGTT).

The fact that HbA1c is subject to non-genetic factors outside the control of healthcare providers should be taken into consideration to prevent arbitrary comparisons such as race, ethnicity, age, diabetes length and type, patient adherence, and comorbidity. This consideration helps to prevent prediction that reflects race/ethnicity biases in making clinical decisions. With that being said, appropriate data sample selection in a population has to be narrowed down into a racial scale. To deliver high predictive power and promise to transform patient risk stratification without any biases, the selected population should be a single-racial population country like Japan.

ACKNOWLEDGMENTS

The authors also wish to express gratitude to the management of Asia Pacific University of Technology & Innovation (APU) for their support.

Conflict of Interest

The authors involved in the current study does not declare any competing conflict of interest.

Funding and Sponsorship

No fund or sponsorship in any form was obtained from any organization for carrying out this research work.

REFERENCES

1. El-Kassas G, Ziade F. Exploration of the dietary and lifestyle behaviours and weight status and their self-perceptions among health sciences university students in north Lebanon. *BioMed Res Int.* 2016 Jan 1;67-79.
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. *Diab Res Clin Pract.* 2019 Nov 1;157:107843.
3. Goudjerkian T, Jayabalan M. Predicting 30-day hospital readmission for diabetes patients using multilayer perceptron. *Int J Adv Comp Sci Appl.* 2019 Feb 28;10(2).
4. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diabetes Care.* 2019 Jan 1;42(Supplement 1): S13-28.

5. World Health Organization. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. World Health Organization; 2011. World Health Organization. Classification of diabetes mellitus. Geneva: World Health Organization; 2019. License: CC BY-NC-SA 3.0 IGO
6. Beam AL, Kohane IS. Big data and machine learning in health care. *J Am Med Assoc.* 2018;319(13):1317-1318.
7. Pop TL, Namazova L, Mestrovic J, Nigri L, Vural M, Sacco M. The Role of Healthy Lifestyle Promotion, Counselling, and Follow-up in Noncommunicable Diseases Prevention. *J Ped.* 2019; 10(23): 278-281.
8. World Health Organization. (2018). Noncommunicable diseases country profiles 2018. World Health Organization. <https://apps.who.int/iris/handle/10665/274512>. License: CC BY-NC-SA 3.0 IGO
9. Bleich SN, Vercammen KA, Zatz LY, Frelief JM, Ebbeling CB, Peeters A. Interventions to prevent global childhood overweight and obesity: a systematic review. *Lancet Diab Endoc.* 2018 Apr 1;6(4):332-346.
10. Verma S, Hussain ME. Obesity and diabetes: an update. *Diabetes & Metabolic Syndrome: Clin Res Rev.* 2017 Jan 1;11(1):73-9.
11. Al-Nakeeb Y, Lyons M, Dodd LJ, Al-Nuaim A. An investigation into the lifestyle, health habits and risk factors of young adults. *Int J Env Res Pub Heal.* 2015 Apr;12(4):4380-94.
12. Sacks DB. A1C Versus Glucose Testing: A Comparison. *Diab Care.* 2011; 34(2):518-523.
13. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2018. *Diab Care.* 2018 Jan;41(1): S55.
14. Mellitus D. Diagnosis and classification of diabetes mellitus. *Diab Care.* 2005 Jan 1;28(S37): S5-10.
- 15.
16. Lin H, Yi J. Current status of HbA1c biosensors. *Sensors. Diab Care.* 2017 Aug;17(8):76-98.
17. Lim WY, Ma S, Heng D, Tai ES, Khoo CM, Loh TP. Screening for diabetes with HbA1c: Test performance of HbA1c compared to fasting plasma glucose among Chinese, Malay and Indian community residents in Singapore. *Sci Repor.* 2018 Aug 20;8(1):1-9.
18. Bergenstal RM, Gal RL, Connor CG, Gubitosi-Klug R, Kruger D. Racial differences in the relationship of glucose concentrations and haemoglobin A1c levels. *Ann Int Med.* 2017 Jul 18;167(2):95-102.
19. Mohammed M, Khan MB, Bashier EB. Machine learning: algorithms and applications. Crc Press; 2016 Aug 19.
20. Kalina J. Big Data, Biostatistics and Complexity Reduction. *Eur J Dia Care.* 2018;14(2):24-32.
21. Ngiam KY, Khor W. Big data and machine learning algorithms for health-care delivery. *Lancet Onc.* 2019 May 1;20(5):e262-73.
22. FDA. Changes to existing medical software policies resulting from section 3060 of the 21st Century Cures Act—Draft Guidance for Industry and Food and Drug Administration Staff.
23. Gepperth A, Hammer B. Incremental learning algorithms and applications. *Lancet Onc.* 2017; 47(10): 1261-1266
24. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *New Eng J Med.* 2018 Mar 15;378(11):981.
25. Stern AD, Price WN. Regulatory oversight, causal inference, and safe and effective health care machine learning. *Biostat.* 2020 Apr 1;21(2):363-7.
26. Rivera-Caravaca JM, Roldan V, Esteve-Pastor MA, Valdes M, Vicente V, Lip GY, Marin F. Long-Term Stroke Risk Prediction in Patients With Atrial Fibrillation: Comparison of the ABC-Stroke and
27. Shah NH, Liu D, Lee C. CHA₂DS₂-VAS_c Scores. *J Am Heart Ass.* 2017 Jul 20;6(7):e006490.
28. Karmali KN, Lloyd-Jones DM. Implementing cardiovascular risk prediction in clinical practice: the future is now. *J Am Heart Ass.* 2017; 9(10): e005672.
29. Arcadu F, Benmansour F, Maunz A, Willis J, Haskovo Z, Prunotto M. Deep-learning algorithm predicts diabetic retinopathy progression in individual patients. *J Dig Med.* 2019 Sep 20;2(1):1-9.
30. Ruggles DR, Freyman RL, Oxenham AJ. Influence of musical training on understanding voiced and whispered speech in noise. *PloS one.* 2014 Jan 28;9(1):e86980.
31. Del Parigi A, Tang W, Liu D, Lee C, Pratley R. Machine Learning to Identify Predictors of Glycemic Control in Type 2 Diabetes: An Analysis of Target HbA1c Reduction Using Empagliflozin/Linagliptin Data. *Pharm Med.* 2019;33(3):209-217.
32. Grigory Sidorenkov MD, van Boven PharmD JF. Predicting short-and long-term glycated haemoglobin response after insulin initiation in patients with type 2 diabetes mellitus using machine-learning algorithms. *Pharm Med.* 2018; 24(7):412-417.