Mapana – Journal of Sciences
2024, Vol. 23, Special Issue 3, 113-120
ISSN 0975-3303 | https://doi.org/10.12723/mjs.sp3.7

Prediction of Type2 Diabetes using Insulin DNA Sequence

Aswathi Sasidharan* & Arulkumar N†

Abstract

This research paper addresses the challenge of objectively evaluating diverse biological characteristics through the classification of DNA sequences. Identifying DNA sequences in genomics research can aid in discovering novel protein activities, such as insulin, which regulates blood sugar levels in the human body. Diabetes, a prevalent chronic illness, is linked to changes in the insulin gene sequence. The study aims to develop a machine-learning model to categorize the insulin gene's DNA sequence and identify type 2 diabetes based on this transformation. The model's performance will be compared to existing machine-learning models. Additionally, the research seeks to identify unique gene variants of the insulin protein associated with diabetes prognosis and investigate the risk factors associated with these gene variants.

Keywords: Machine Learning, Type2 Diabetes, Machine Learning, DNA Sequence, Model Construction, BLAST, and AUGUSTUS.

1. Introduction

Metabolism is how our bodies turn the food and water we consume into energy. This chemical reaction happens inside a living organism. This process produces energy by combining the caloric content of food and beverages with oxygen [1]. Nevertheless, a metabolic disorder

* Ph.D. Scholar, Department of Computer Science, CHRIST (Deemed to be University). Bangalore - 560029, INDIA.
† Assistant Professor, Department of Statistics and Data Science, CHRIST(Deemed to be University). Bangalore - 560029, INDIA.
Corresponding Author: arul.kumar@christuniversity.in

results when these chemical processes are impeded, maybe because the liver or pancreas is ill or not functioning regularly. Very high blood glucose or blood sugar levels may result in diabetes. Blood glucose, derived from food, is the primary energy source. Insulin is a protein generated by the pancreas that allows glucose to enter cells for use as energy. Sometimes, our body either does not create insulin or poorly utilizes it [2]. Diabetes is the effect of this. Diabetes is a chronic illness that may have severe unfavorable effects if not adequately diagnosed and controlled. Each kind of diabetes has its own set of causes, but they all share the problem of high glucose levels. As therapeutic alternatives, insulin, and medicines are employed. A healthy lifestyle may help avoid some types of diabetes [3,4].

1.1 Types of Diabetes

With autoimmune disorders like type 1, the immune system can attack itself. This results in the death of pancreatic cells that produce insulin. Around 10% of people with diabetes are diagnosed with type 1 diabetes [5,6]. Children and teenagers are often diagnosed (but can develop at any age). Often, diabetes was referred to as "juvenile diabetes." Patients with diabetes type 1 necessity take insulin regularly. Because of this, it is often referred to as insulin-dependent diabetes. When insulin synthesis in the body is insufficient or cells no longer react normally to insulin, this is known as type 2 diabetes.

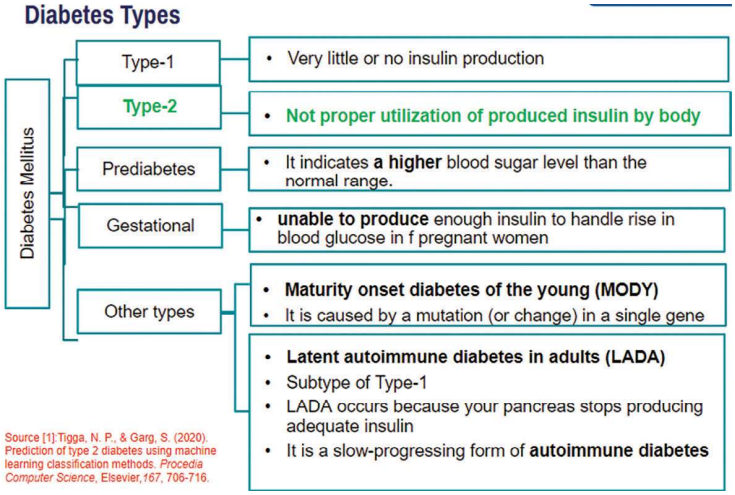

Figure 1: Types of Diabetes Mellitus [14]

The most prevalent kind of diabetes. Around 95% of diabetes patients have Type 2 diabetes. Those in their middle years and older are often afflicted. Insulin-resistant diabetes and adult-onset diabetes are two additional terms for Type 2 diabetes [7, 8]. This condition comes before type 2 diabetes. Despite the blood glucose levels being higher than usual, Type 2 diabetes cannot be diagnosed based only on these measurements. One kind of diabetes that may affect pregnant women is gestational diabetes. Often, gestational diabetes disappears after pregnancy. Type 2 diabetes is more likely to develop in those with gestational diabetes.

2 Literature Review

Ref No.	Title	Findings
1	Awotunde, J. B. et al. [9]	• This document revises the World Health Organization's classification of diabetes from 1999. (WHO). • Type 1 and Type 2 diabetes (T2DM), according to the current classification system, are the two most frequent types of illness (T2DM).
2	Dias et al. [10]	• Concentrate on creative methodologies for specialized clinical genomics applications, Eg. variant categorization. • Identified barriers related to AI in clinical diagnosis.
3	Guo et al. [11]	• GDL (deep genome learning) developed a method using deep neural networks to examine the relationship between genetic variations and traits.
4	Lai, Hang, et al. [12]	• Diabetes has been predicted using decision trees, random forests, and neural networks. • Using five-fold cross-validation, the models were assessed.
5	Akbarzadeh, Mahdi, et al. [13]	• Assesses several machine learning-based classification algorithms for detecting metabolic syndrome and locating pertinent genetic and environmental risk variables. • The models used GCKR gene polymorphisms and clinical and demographic data.

3. The Proposed Model

The proposed research intends to categorize insulin sequences from diverse human specimens as type 2 diabetic or non-type two diabetics to develop a predictive algorithm for type 2 diabetes. Employing machine learning techniques to do this may be achievable. With the collected data and algorithms, it is feasible to anticipate type 2 diabetes. In this case, the genome-wide expression of every gene is examined simultaneously. A mutation in the insulin gene's sequence causes an uneven synthesis of insulin, which causes diabetes.

Nucleotides are DNA's building blocks. Each nucleotide unit is composed of three subunits. Examples of nitrogenous bases include a sugar with at least one phosphate group, another sugar, and still another sugar. Adenine, guanine, cytosine, and thymine are the nitrogenous bases found in DNA. Nucleotide sequence variations result in biological variability in all living things, including people. The arrangement of these nitrogenous bases encrypts biological information. Genes from other closely related genes must be translated and transcribed to produce proteins. A specific DNA region is transcribed into a transitory mRNA molecule during transcription. Base pairs are the primary means of communication between nucleic acids like DNA and RNA. Translation turns mRNA into protein. Gene expression involves these actions. The DNA and RNA are shown in Figure 2.

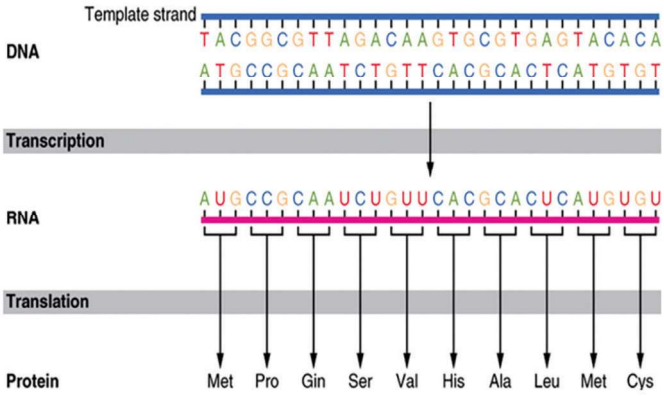

Figure 2: DNA and RNA

Beta cells in the pancreas generate insulin to control blood glucose levels. In a healthy body, beta cells release insulin and increase hormone synthesis in response to rising blood glucose levels. However, in some circumstances, such as type 2 diabetes, beta cells cannot produce enough insulin to manage blood glucose levels. Diabetes results from insulin gene sequence changes that disrupt insulin production.

4. Details of the Study

The human body produces insulin to manage blood sugar levels. Diabetes, one of the most prevalent chronic conditions, may have significant long-term consequences if not adequately identified and controlled. Diabetes Mellitus occurs from alterations to the insulin gene's sequence. Classifying DNA sequences presents a substantial problem for scientists who want to examine vast biological data and find various biological properties objectively. Most of this field's research utilizes electronic health data, which has been helpful but has considerable limitations. There are inconsistencies across websites. Several databases may yield differing estimations of the relationship between biological parameters and electronic health data for a given study design.

This paper introduces a unique Machine Learning model for categorizing the DNA sequence of the insulin gene and predicting the onset of type 2 diabetes. This model yields more accurate results using genetic data than models incorporating biological factors. Because that genotype-based knowledge of drug-metabolizing enzymes may affect the response to pharmacological treatment, this notion may be broadened to include precision medicine. The proposed methodology is given in Figure 3.

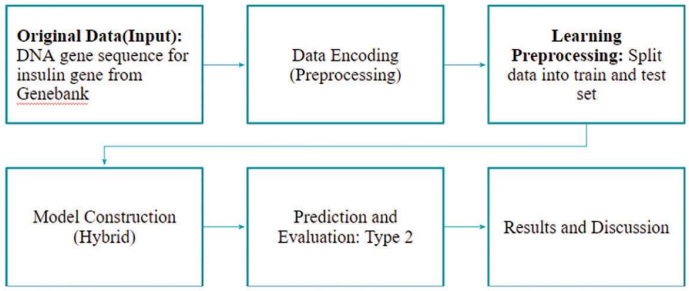

Figure 3: Proposed Methodology

The following are the different stages of the proposed methodology.

- ***Input:*** *From Genebank, the DNA code for the insulin gene*
- ***Data-Encoding (Pre-processing):*** *Pre-process the data using one hot encoding*
- ***Learning Pre-processing:*** *Split data into train and test set*
- ***Model-Construction:*** *Build the model on the training set*
- ***Data-Evaluation:*** *Prediction of type2 diabetes and validation of the model*

5. Tools Identified for Research Work Implementation

Python is a dynamically semantic, object-oriented, high-level programming language. It is a translation of a language. Python's concise and straightforward syntax enhances readability, minimizing software maintenance costs. Python allows modules and packages, increasing programs' modularity and reusability.

5.1 BLAST Basic Local Alignment Search Tool (BLAST)

BLAST finds local sequence similarities. The app evaluates the statistical significance of nucleotide or protein sequence matches to sequence databases.

5.2 Augustus

An expanded hidden Markov model-based gene prediction programme for eukaryotic genomic sequences describes the sequence and gene structure probabilistically.

6. Conclusion

This research found that using machine learning to genetic data might be beneficial for predicting type 2 diabetes. This strategy may give a more thorough analysis of massive data sets and a deeper understanding of human health and disease. In addition, the classification of DNA sequences may aid researchers in understanding protein activities in genomic research and developing innovative treatments and methods for preventing type 2 diabetes. Using this innovative technology, researchers may be able to diagnose type 2 diabetes with more accuracy and uncover novel therapies for the disease.

References

[1] Gulcin, İlhami. "Antioxidants and antioxidant methods: An updated overview." *Archives of toxicology* 94, no. 3 (2020): 651-715.

[2] Liu, Jun-Li, Irina Segovia, Xiao-Lin Yuan, and Zu-hua Gao. "Controversial roles of gut microbiota-derived short-chain fatty acids (SCFAs) on pancreatic β-cell growth and insulin secretion." *International journal of molecular sciences* 21, no. 3 (2020): 910.

[3] Neelakandan, S., J. Rene Beulah, L. Prathiba, G. L. N. Murthy, E. Fantin Irudaya Raj, and N. Arulkumar. "Blockchain with deep learning-enabled secure healthcare data transmission and diagnostic model." *International Journal of Modeling, Simulation, and Scientific Computing* 13, no. 04 (2022): 2241006.

[4] Jaishankar, B., Santosh Vishwakarma, Prakash Mohan, Aditya Kumar Singh Pundir, Ibrahim Patel, and N. Arulkumar. "Blockchain for Securing Healthcare Data Using Squirrel Search Optimization Algorithm." *Intelligent Automation & Soft Computing* 32, no. 3 (2022).

[5] Rathod, Sanjay. "Novel insights into the immunotherapy-based treatment strategy for autoimmune type 1 diabetes." *Diabetology* 3, no. 1 (2022): 79-96.

[6] Manimaran, Aridoss, Dhasarathan Chandramohan, S. G. Shrinivas, and N. Arulkumar. "A comprehensive novel model for network speech anomaly detection system using deep learning approach." *International Journal of Speech Technology* 23 (2020): 305-313.

[7] Artasensi, Angelica, Alessandro Pedretti, Giulio Vistoli, and Laura Fumagalli. "Type 2 diabetes mellitus: a review of multi-target drugs." *Molecules* 25, no. 8 (2020): 1987.

[8] Satish Kumar, T., S. Jothilakshmi, Batholomew C. James, M. Prakash, N. Arulkumar, and C. Rekha. "HHO-based vector quantization technique for biomedical image compression in cloud computing." *International Journal of Image and Graphics* (2021): 2240008.

[9] Awotunde, J. B., Ayo, F. E., Jimoh, R. G., Ogundokun, R. O., Matiluko, O. E., Oladipo, I. D., & Abdulraheem, M. (2021). Prediction and classification of diabetes mellitus using genomic data. In *Intelligent IoT systems in personalized health care* (pp. 235-292). Academic Press.

[10] Dias, Raquel, and Ali Torkamani. "Artificial intelligence in clinical and genomic diagnostics." *Genome medicine* 11.1 (2019): 1-12.

[11] Guo, Yang, Xuequn Shang, and Zhanhuai Li. "Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer." *Neurocomputing* 324 (2019): 20-30.

[12] Lai, Hang, et al. "Predictive models for diabetes mellitus using machine learning techniques." *BMC endocrine disorders* 19.1 (2019): 1-9.

[13] Akbarzadeh, Mahdi, et al. "Evaluating machine learning-powered classification algorithms which utilize variants in the GCKR gene to predict metabolic syndrome: Tehran Cardio-metabolic Genetics Study." *Journal of translational medicine* 20.1 (2022): 1-12.

[14] Tigga, Neha Prerna, and Shruti Garg. "Prediction of type 2 diabetes using machine learning classification methods." *Procedia Computer Science* 167 (2020): 706-716.