Application of Genetic Algorithms for Medical Diagnosis of Diabetes Mellitus

Khushali Tyagi¹, Deepak Kumar¹* and Richa Gupta²

¹Department of Applied Sciences, Manav Rachna International Institute of Research and Studies, Faridabad, India;
²Global Institute of Technology and Management, Gurgaon, India

Abstract: The system of glucose-insulin control and associated problems in diabetes mellitus were studied by mathematical modeling. It is a helpful theoretical tool for understanding the basic concepts of numerous distinct medical and biological functions. It delves into the various risk factors contributing to the onset of diabetes, such as sedentary lifestyle, obesity, family history, viruses, and increasing age. The study emphasizes the importance of mathematical models in understanding the dynamic characteristics of biological systems. The study emphasizes the increasing prevalence of diabetes, especially in India, where urbanization and lifestyle changes contribute to the rising incidence. The present investigation describes the use of John Holland’s evolutionary computing approach and the Genetic Algorithm (GA) in diabetes mellitus. The Genetic Algorithm is applied to address issues related to diabetes, offering a generic solution and utilizing MATLAB’s Genetic Algorithm tool. The Mathematical Model provides differential equations representing glucose and insulin concentrations in the blood. The results represent testing outcomes for normal, prediabetic, and diabetic individuals, optimized with Genetic Algorithm showcased through fitness value plots. The conclusion highlights the effectiveness of Genetic Algorithm as an optimization tool in predicting optimal samples for diabetes diagnosis. The paper encourages the use of heuristic algorithms, such as Genetic Algorithms, to address complex challenges in the field of diabetes research. Future scope includes further exploration of biomathematics and Genetic Algorithm applications for enhanced understanding and management of diabetes mellitus. It is critical for people with diabetes to consistently check their blood glucose levels and follow their treatment plan.

Introduction

Diabetes Mellitus, a chronic metabolic disorder characterized by elevated blood glucose levels, poses a significant global health challenge (Jaiswal et al., 2023). Timely and accurate diagnosis is paramount for effective management and prevention of complications. In recent years, the integration of advanced computational techniques has revolutionized medical diagnostics, and one such promising avenue is the application of Genetic Algorithms (GAs).

Genetic Algorithms, inspired by the principles of natural selection and genetics, are optimization algorithms that simulate the process of natural evolution. Their adaptability and ability to find optimal solutions in complex and dynamic environments make them a compelling tool for addressing the intricate challenges in medical diagnosis, particularly in the context of diabetes. This paper explores the burgeoning field of applying Genetic Algorithms to the medical diagnosis of Diabetes Mellitus. By leveraging the power of evolutionary computing, researchers aim to enhance the efficiency and accuracy of diagnostic processes, leading to improved patient outcomes. This exploration involves the integration of genetic algorithms with medical datasets,
leveraging their capability to evolve and adapt over successive generations to identify optimal sets of features for diagnosis.

As we delve into the intricacies of this innovative approach, it becomes evident that the synergy between genetic algorithms and medical diagnostics has the potential to revolutionize the way we identify, classify, and predict diabetes. This paper will delve into the theoretical foundations, methodologies, and real-world applications of Genetic Algorithms in the medical diagnosis of Diabetes Mellitus, shedding light on their transformative impact on healthcare practices.

Diabetes is a long-term condition characterized by recurrent hyperglycemia. The mathematical modeling of the glucose-insulin regulation system and its complications is expanding all the time, revealing new details about the underlying mechanisms. Mathematical models are essential for studying the dynamic characteristics and mechanisms that drive a set of complex biological systems. Various mathematical models focusing on various aspects of diabetes mellitus have been developed (Al Ali et al., 2022; Balasubramanian et al., 2023). Although the precise origin of diabetes mellitus is not entirely understood, there are a number of established risk factors like Sedentary lifestyle, Obesity, Family history, Virus, Increasing age etc.

The study of biology and medicine is one of the most recent areas of mathematical research. Biomathematics is the term used to describe the contributions of mathematics to biology. There are chances for attractive study in these domains for mathematicians at all levels of interdisciplinary collaboration.

Biomathematics is a rapidly growing topic that has developed into an important aspect of mathematics for the progress of biology and medical science. Biologists can now understand the complexity of various biological phenomena (Pawar et al., 2023). It underlies the construction of a new generation of models, where ideas from various disciplines of science and technology are incorporated to describe many biological processes in a way that predictions can be made. This paper summarises the history and background of the work done in the field of modelling diabetes mellitus by different authors and provides basic knowledge of the work (Sarkar et al., 2023). It gives the motivation for future work. WHO, American Diabetic Association and other reports are studied to collect and obtain correct information for further improvements. Chronic diseases are an important area of research due to the increasing number of people affected worldwide (Kumar et al., 2021; Rosli et al., 2020). When an individual has diabetes, the body doesn't create sufficient insulin or can't involve its insulin as it ought to. This causes the development of sugar in the blood, prompting difficulties like coronary heart disease, stroke, and neurological infection (Gupta and Kumar, 2017). A lack of blood circulation prompts amputations, visual impairment, kidney disappointment, and nerve harm. Diabetes must be diagnosed to be effectively treated; otherwise, the patient may have everlasting consequences that increase treatment expenses and expose them to numerous hazards, as was already noted.

In India, diabetes is a growing public health concern, with a sizable section of the population at risk of having the disease at some point in the future. In India, those who reside in urban areas and large cities are more likely than ever to get diabetes (Sur et al., 2023; Biswas et al., 2023; Roy et al., 2023). To some extent, this is predicted because metropolitan regions are developing a style of life that might change a person's weight profile (BMI). A risk factor for diabetes is a higher BMI. Further research is required to fully understand how and why type 2 diabetes is occurring in rural India, where incidences are also on the rise (Acharya et al., 2023). Keep researching to learn more about diabetes in India, including its prevalence a few very common causes, and that's only the beginning. Approximately 77 million adults in India have diabetes. According to research, by 2045, 134 million people are predicted to be living in this area. In general, women are more likely than men to have diabetes. However, this risk will decrease as both sexes become older (beyond the age of 80). Despite the increasing prevalence of diabetes, experts estimate that 57% of cases remain unidentified. This is especially concerning since those who don't use prescription medication to regulate their blood sugar run a higher risk of developing catastrophic consequences. As the country experiences an
urbanisation phase, diabetes rates are rising, suggesting that more people are relocating to larger metropolitan areas for work and business opportunities. Urbanized areas and communities alter the manner of life, making it more sedentary and riskier for diabetes. The number of diabetics is growing as a result of population growth, ageing populations, urbanisation, increasing prevalence of obesity, and sedentary lifestyles. To enable wise preparedness and allocation of resources, it is essential to assess the prevalence of diabetes and the number of people affected by it, both now and in the future (Wild et al., 2004).

Materials and Methods

This advent of a new frontier is the discipline of Biology and Medical sciences in mathematics area. The study of biomathematics utilizes math in the field of biology. Biomathematics is a scientific research discipline that adopts multiple mathematical models to understand various life science issues and medical practice. The final goal of biomathematics lies in developing mathematical models of biological processes with a wealth of applied mathematical topics and techniques from different mathematical fields. Mathematician can bring his/her research interest into this area from pure mathematics to more practical interdisciplinary forms. Indeed, nowadays, biomathematics is a developing area and acts as a key subdivision of mathematics that influences the practices of modern technology and science. Biologists are now able to understand the complexity of various biological phenomena. It underlies the construction of a new generation of models, where ideas from various disciplines of science and technology are incorporated to describe many biological processes so that predictions can be made.

The mathematical technique finds the best possible solution to a real-life problem under particular constraints and availability. In an optimization problem, the model is familiar with the required output and the task is to find the input from this output.

Maximum real-life optimization problems have different types of variables. Many objectives contradict each other. The output change is not proportional to the input change, lack of continuity, and a region in which curvature extends inwards. It is not easy to find a globally optimal solution in such a wide design space in a limited interval of time (Pramanik, 2018). A solution to this type of problem using several methods may be ineffective or cost-effective. In such circumstances, the methods of choice for solving many hard combinatorial problems, such as simulated annealing, evolutionary algorithms, and hill-climbing, may be applied. Evolutionary algorithms have the capability of dealing with various levels of difficulty (several-objective function, non-linearity, etc.) and may also be used in conjunction with any existing local search or various approaches (Park and Nam, 2023).

The genetic algorithm has a random base and is a classic evolutionary algorithm (Santhanam and Padmavathi, 2015). In Genetic Algorithm, new solutions are generated by applying random modifications to the current ones to eventually find a solution (Sapna et al., 2012; Alharbi et al., Nandhini and Dharmarajan, 2023). GA is sometimes referred to as Simple GA since it is simpler than other EAs (SGA).

Genetic Algorithm is rooted in Darwin's theory of evolution and involves a gradual, cumulative process that introduces small, often imperceptible modifications to solutions until the best option is identified. The algorithm continues to make these incremental changes until an optimal solution is found. The total number of solutions equals the population size (pop size) of the population on which GA operates. Every response is referred to as being distinct. Every distinct solution contains a chromosome. The chromosome is a set of traits (parameters) defining a person. Each chromosome may include a set of genes. It is used for glucose normalization in diabetic patients (Jasim and Al Mashhadany, 2021; Gamboa et al., 2022; Medhi et al., 2023).

Description of the Genetic Algorithm

The genetic algorithm that the late John Holland (known as father of Genetic Algorithm) developed in the early70s is the one I am talking about. These can be defined as those that are based on population models and artificial neural networks. A straight line that shows how a gas is regulated is briefly explained here (McCall, 2005; Babaei, 2013; Parlikar and Dahe, 2019):

Initialization:

- Random selection/creation of the initial population (set of solutions) results in the commencement of the algorithm.

Selection:

- Identify those with the sounding solutions in a population.

Fitness evaluation:

- Every value is related to the corresponding solution which is measured on the accuracy level with the optimal solution to the given problem.

Crossover:

- The gene flow from one group to another happens as combinations of solutions occur while a mating pool is in use.
Mutation:
The process is implemented by developing child organisms from one parent (hence open mutation) and searching for the best solution (optimal solution) is also included.

Elitism:
The term elite is assigned to the existence of a super limited number of the best solutions.

Using a genetic algorithm, here is a generic solution to the issue

The Genetic Algorithm tool, or GUI, can be used to tackle the issues (Graphical User Interface). It enables the use of GA without having to use the command line. Entering “Gatool” at the MATLAB command prompt will launch the genetic Algorithm tool. This example demonstrates how to use the Global Optimization Toolbox’s Genetic Algorithm (GA) solver to resolve a problem.

Literature Review of Genetic Algorithm for Diabetes Mellitus

When the relative magnitudes of data objects are more relevant than their accurate values, the Genetic Algorithm is useful in capturing related patterns in the data. This algorithm will improve by allowing users to readily analyze data from a vast database quickly and with more accuracy (Kovalan et al., 2014).

In this study, A new approach has been developed to help doctors assess their patient’s condition and choose the optimum pharmacological treatment option. It extracts the optimal test sequence evaluation mechanism.

A system for predicting diabetes that combines a genetic algorithm and a Gaussian radial basis function network (GAGRBFN) is suggested. The suggested system was trained and evaluated using the PIMA Indian Diabetes Dataset, and classification accuracy and error were used to gauge the system’s efficacy. The simulation results show that the suggested system is more accurate and less error-prone than the sophisticated methods presented (Yadav et al., 2022).

Mathematical Model

Biomathematics is a branch of mathematics with many applications in biology and medicine. Biomathematics’ major goal is to use various applied
mathematical tools and methodologies to develop mathematical models of biological processes. We consider the concentrations of glucose and insulin in the blood. The differential equations are defined as the excess concentration of glucose and the excess concentration of insulin over their respective equilibrium values at time t.

\[
\frac{dG}{dt} = -m_1G - m_2I + g(t), \quad (1)
\]

\[
\frac{dI}{dt} = -m_3I - m_4G + (i(t), \quad (2)
\]

In the given context, "G" represents the basal concentration of glucose and "I" represents the basal concentration of insulin.

When m1, m2, m3 and m4 > 0 for the following reasons:
1. Excess glucose tends to be absorbed by the liver and tissues, causing m1>0.
2. If there is extra insulin, it aids in the tissues’ ability to metabolise glucose such that m2>0.
3. Extra insulin tends to vanish if there is any, causing m3>0.
4. The pancreas is stimulated to release insulin if there is too much glucose, preventing m4>0.

By eliminating, step-by-step, G, I from (1) and (2), we get

\[
\alpha^2 \omega^2 G + 2\alpha \frac{dG}{dt} + \omega^2 G = \dot{S}_1(t), \quad (3)
\]

\[
\alpha^2 \omega^2 I + 2\alpha \frac{dI}{dt} + \omega^2 I = \dot{S}_2(t), \quad (4)
\]

Where
\[
2\alpha = m_1 + m_3, \quad (5)
\]

\[
\alpha^2 = m_1m_3 + m_2m_4, \quad (6)
\]

If \( \dot{S}_1(t) = e^{-\alpha t} \sin \omega t, \dot{S}_2(t) = 0, \) then (3) and (4) give the general solutions

\[
G = e^{-\alpha t} (X \cos \omega t + Y \sin \omega t) + \frac{e^{-\alpha t}}{\omega} B \sin \omega t \quad (9)
\]

\[
I = e^{-\alpha t} (M \cos \omega t + N \sin \omega t) \quad (10)
\]

If t=0, then G=0, so that X=0. Also, integrating (3), we get

\[
\int_0^t e^{-\alpha t} \sin \omega t \ dt = \frac{e^{-\alpha t}}{\omega} B \sin \omega t - \sin \omega t + \frac{B}{\omega^2} \quad (11)
\]

So that

\[
G = \frac{e^{-\alpha t}}{\omega} B \sin \omega t \quad (13)
\]

### Table 2. Standard Blood sugar levels in diagnosing diabetes (Sandhya and Kumar, 2019)

<table>
<thead>
<tr>
<th>Plasma Glucose test</th>
<th>Normal</th>
<th>Prediabetes</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting</td>
<td>Below 100mg/dl</td>
<td>100 to 125 mg/dl</td>
<td>126 mg/dl or more</td>
</tr>
<tr>
<td>2 hour post-prandial</td>
<td>Below 140 mg/dl</td>
<td>140 to 199 mg/dl</td>
<td>200/m/d or more</td>
</tr>
</tbody>
</table>

By eliminating, step-by-step, G, I from (1) and (2), we get

\[
\frac{d^2 G}{dt^2} + 2\alpha \frac{dG}{dt} + \omega^2 G = \dot{S}_1(t) \quad (7)
\]

\[
\frac{d^2 I}{dt^2} + 2\alpha \frac{dI}{dt} + \omega^2 I = \dot{S}_2(t) \quad (8)
\]

Result and Discussion

Applying the Genetic algorithm of the equation (13) and find the results. Table 2 shows the testing results for a Normal, Prediabetic and Diabetic person:

To achieve optimal results, Genetic algorithm is used for Normal Person

Figure 3a plot is displayed, which represents the graphical representation of the optimal or most favourable value and mean value (100.569) of fasting for a normal person and over the iteration of 51 iterations. Figure 3b shows the plot of the best value
and mean value (129.202) of 2-hour postprandial for a normal person and over the iteration of 59 iterations.

Figure 3a. The fitness value with the highest score after 51 iterations of fasting for a Normal Person

Figure 3b. The fitness value with the highest score after 59 iterations of 2 hours postprandial for Normal Person.

To achieve optimal results, Genetic algorithm is used for Prediabetic Persons

Figure 4a plot is displayed, which represents the graphical representation of the optimal or most favourable value and mean value (120.683) of fasting for a Prediabetic person and over the iteration of 51 iterations. Figure 4b shows the plot of the best value and mean value (176.184) of 2 hour postprandial for a Prediabetic person and over the iteration of 52 iterations.

Figure 4a. The fitness value with the highest score after 51 iterations of fasting for a Prediabetic Person

Figure 4b. The fitness value with the highest score after 52 iterations of 2 hour postprandial for Prediabetic Person.

To achieve optimal results, Genetic algorithm is used for Diabetic Person

Figure 5a plot is displayed, which represents the graphical representation of the optimal or most favourable value and mean value (150.854) of fasting for a Diabetic person and over the iteration of 51 iterations. Figure 5b shows the plot of the best value and mean value (205.548) of 2 hour postprandial for a Diabetic person and over the iteration of 59 iterations. The above results denote the fitness value with the highest score and mean fitness values, the points above them, however, represent the average fitness values throughout generations.
### Table 3. Fitness chart of experiments.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Normal Person</th>
<th>Prediabetic Person</th>
<th>Diabetic Person</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fasting- Mean fitness</td>
<td>2 hour post-prandial Mean fitness</td>
<td>Fasting- Mean fitness</td>
</tr>
<tr>
<td>1</td>
<td>119</td>
<td>230</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>190</td>
<td>137</td>
</tr>
<tr>
<td>3</td>
<td>107</td>
<td>161</td>
<td>129</td>
</tr>
<tr>
<td>4</td>
<td>106</td>
<td>150</td>
<td>126</td>
</tr>
<tr>
<td>5</td>
<td>105</td>
<td>140</td>
<td>122</td>
</tr>
<tr>
<td>6</td>
<td>104</td>
<td>135</td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>132</td>
<td>121</td>
</tr>
<tr>
<td>8</td>
<td>101</td>
<td>131</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>101</td>
<td>130</td>
<td>121</td>
</tr>
<tr>
<td>10</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>11</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>12</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>13</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>14</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>15</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>16</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>17</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>18</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>19</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>20</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>21</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>22</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>23</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>24</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>25</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>26</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>27</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>28</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>29</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>30</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>31</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>32</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>33</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>34</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>35</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>36</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
<tr>
<td>37</td>
<td>100.56</td>
<td>129.2</td>
<td>120.683</td>
</tr>
</tbody>
</table>
We have a set of feasible solutions that can be applicable where needed. This is remarkable benefited of genetic algorithm. Using the Genetic Algorithm, it is optimized. The normal person's diabetes mellitus level is 100.56 mg/dL during fasting and 129.2 mg/dL two hours after meals. The Prediabetic Person's diabetes mellitus level is 120.68 mg/dL during fasting and 176.18 mg/dL two hours after meals. The diabetic person's diabetes mellitus level is 150.85 mg/dL during fasting and 205.54 mg/dL two hours after meals. The primary indicator of diabetes mellitus is elevated blood glucose levels. Fasting

Blood Glucose and Postprandial Blood Glucose are two important measurements used to monitor blood sugar levels in individuals, particularly those with diabetes.

**Conclusion**

There is an absence of research validating the plants identified throughout this survey's possible anti-diabetic effects. Monitoring both fasting and post-prandial levels is essential for effective diabetes management. Products made from plants have the potential to be effective in
treatning diabetes since they have fewer adverse effects. The efficacy of the Genetic algorithm as an optimization tool in determining optimal samples for prediction has indeed been established in this work. Using a suitable objective function and associated fitness values, the data obtained from biostatistics has been improved. To sum up, heuristic algorithms like the Genetic Algorithm may be used to tackle several problems. It will provide us with strong instruments to tackle really difficult challenges easily. It is important for individuals with diabetes to regularly monitor their blood glucose levels and adhere to their treatment. One possible area of study might relate to diet scheduling for diabetes patients with additional medical complications.

Conflict of Interest
The authors declare no conflict of interest.

References


DOI: https://doi.org/10.52756/ijerr.2024.v37spl.001


How to cite this Article:
DOI : https://doi.org/10.52756/ijerr.2024.v37spl.001

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.