Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: G

Interdisciplinary

Diabetes in Human Body

Practicing Clinicians from India

Highlights

Machine Learning Technique

P vs NP: P is Equal to NP: Desired

Discovering Thoughts, Inventing Future

VOLUME 21 ISSUE 3 VERSION 1.0

Global Journal of Computer Science and Technology: G
Interdisciplinary

Global Journal of Computer Science and Technology: G
Interdisciplinary

Volume 21 Issue 3 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

Editorial Board

Global Journal of Computer Science and Technology

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajesh Kumar Rolen

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MIcrosoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

Contents of the Issue

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 3 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

P vs NP: P is Equal to NP: Desired Proof

By Zulfia A. Chotchaeva
Moscow State University

Abstract- Computations and computational complexity are fundamental for mathematics and all computer science, including web load time, cryptography (cryptocurrency mining), cybersecurity, artificial intelligence, game theory, multimedia processing, computational physics, biology (for instance, in protein structure prediction), chemistry, and the P vs. NP problem that has been singled out as one of the most challenging open problems in computer science and has great importance as this would essentially solve all the algorithmic problems that we have today if the problem is solved, but the existing complexity is deprecated and does not solve complex computations of tasks that appear in the new digital age as efficiently as it needs. Therefore, we need to realize a new complexity to solve these tasks more rapidly and easily. This paper presents proof of the equality of P and NP complexity classes when the NP problem is not harder to compute than to verify in polynomial time if we forget recursion that takes exponential running time and goes to regress only (every problem in NP can be solved in exponential time, and so it is recursive, this is a key concept that exists, but recursion does not solve the NP problems efficiently). The paper's goal is to prove the existence of an algorithm solving the NP task in polynomial running time. We get the desired reduction of the exponential problem to the polynomial problem that takes O(log n) complexity.

Keywords: P vs. NP, P=NP, computational complexity, NP-complete problems, exponential running time.

GJCST-G Classification: I.2.8

Strictly as per the compliance and regulations of:

P vs NP: P is Equal to NP: Desired Proof

Zulfia A. Chotchaeva[1]

Abstract- Computations and computational complexity are fundamental for mathematics and all computer science, including web load time, cryptography (cryptocurrency mining), cybersecurity, artificial intelligence, game theory, multimedia processing, computational physics, biology (for instance, in protein structure prediction), chemistry, and the P vs. NP problem that has been singled out as one of the most challenging open problems in computer science and has great importance as this would essentially solve all the algorithmic problems that we have today if the problem is solved, but the existing complexity is deprecated and does not solve complex computations of tasks that appear in the new digital age as efficiently as it needs. Therefore, we need to realize a new complexity to solve these tasks more rapidly and easily. This paper presents proof of the equality of P and NP complexity classes when the NP problem is not harder to compute than to verify in polynomial time if we forget recursion that takes exponential running time and goes to regress only (every problem in NP can be solved in exponential time, and so it is recursive, this is a key concept that exists, but recursion does not solve the NP problems efficiently). The paper's goal is to prove the existence of an algorithm solving the NP task in polynomial running time. We get the desired reduction of the exponential problem to the polynomial problem that takes O(log n) complexity.

Keywords: P vs. NP, P=NP, computational complexity, NP-complete problems, exponential running time.

I. Introduction

Another mention of the underlying problem occurred in a … letter written by Kurt Gödel to John von Neumann. Gödel asked whether theorem-proving (now known to be co-NP-complete) could be solved in quadratic or linear time, and pointed out one of the most important consequences – that if so, then the discovery of mathematical proofs could be automated (Wikipedia, 2021).

The P vs. NP problem remains one of the most important problems in computational complexity. Until now, the answer to that problem is mainly "no". And this is accepted by the majority of the scientific world. What is the P versus NP problem, and why should we care? The question is represented as P=?NP. P-class problems take polynomial time to solve a problem (less time), NP-class problems take "non-deterministic" polynomial time to quickly check a problem (more time), therefore, P problems are easier to solve while NP problems are harder. NP-complete problems are the hardest and take more time than P-class problems. If P=NP, we could find solutions to search problems as easily as checking since a solution for any NP-class problem can be recast into a solution for any other problem of this class. Thus, finding the efficient algorithm would prove that P=NP and revolutionize (completely turn) many fields in mathematics and computer science. "The development of mathematics in the direction of greater exactness has – as well known – led to large tracts of it becoming formalized so that proofs can be carried out according to the few mechanical rules (Gȍdel, 1931)." "Perhaps in most cases where we seek in vain the answer to the question, the cause of the failure, lies in the fact that problems are simpler and easier than the one in hand have been either not at all or incompletely solved (Hilbert, 2000)." "Do NP-complete languages exist? It may not be clear that NP should process a language that is as hard as any other language in the class. However, this does turn out to be the case (Arora and Barak, 2009)." All previous attempts to solve the problem did not lead to the desired solution. But we declare that the desired solution exists. The paper will get easy proof of the equality of complexity classes P and NP through (with) the new computational complexity that takes polynomial running time and completely rearranges these complexity classes (we will get an exponential-time reduction to polynomial time using a sorted array). The paper intends to prove that the use of logarithmic looping of matrices through a sequence of matrix loops replaces recursive iterating that takes $O(2^n)$ with a completely new and another method (approach) that is more efficient and faster than existing and takes O(log n) complexity instead of $O(2^n)$ when we solve the NP task. (Notice, we will not compare this work and its methods (operators) with what has been done before, because it is so different from everything that already exists that it simply makes it impossible, for instance, like the Boolean satisfiability problem (SAT), the Cook-Levin theorem, the Curry-Howard isomorphism, the Davis-Putnam algorithm, the Davis-Putnam-Logemann-Loveland procedure, the Karp-Lipton theorem, and others that are conversions of the listed, that is, have nothing in common what lies at the basis of these approaches, except for the fractional differentiation, but it does not rely on old, previously

Author: Moscow State University, Krasniy Kurgan 369387, Russia.*
e-mail: lesya.chotchaeva8@gmail.com
**The self-study*

known algorithms, methods, principles, concepts, or models, this light tutorial is completely new and will easily refute the unsolvability, or intractability, of the P vs. NP problem. More precisely, to change the NP, we needed to change the P, i.e., we use a polynomial-time reduction that is the perfect way to provide (get) the reducibility and computability of NP that make the problem of NP the problem of P.)

II. Literature Review

a) Background

The P versus NP problem is a major unsolved problem in computer science. P-complexity is a deterministic polynomial, we consider this complexity class as $O(n^c)$, where the base is variable, and the exponent of the base is constant; and NP-complexity is a nondeterministic polynomial, we consider this complexity class as $O(c^n)$, where the base is constant, and the exponent of the base is variable. The polynomial and exponential time complexities are the most prominently considered and define the complexity of an algorithm. The question is - whether every problem whose solution can be quickly verified in polynomial running time can be solved quickly in polynomial running time too? If NP-complete problems were efficiently solvable, it could advance considerably the solution of other complex problems.

"The precise statement of the P versus NP problem was introduced in 1971 by Stephen Cook in his seminal paper "The complexity of theorem-proving procedures". …Although the P versus NP problem was formally defined in 1971, there were previous inklings of the problems involved, the difficulty of proof, and the potential consequences…The relation between the complexity classes P and NP are studied in computational complexity theory, the part of the theory of computation dealing with the resources required during computation to solve a given problem. The most common resources are time (how many steps it takes to solve a problem) and space…the class P consists of all those decision problems that can be solved on a deterministic sequential machine in an amount of time that is polynomial in the size of the input; the class NP consists of all those decision problems whose positive solutions can be verified in polynomial time … on a non-deterministic machine. …To attack the P=NP question, the concept of NP-completeness is very useful. … NP-complete problems are a set of problems to each of which any other NP-problem can be reduced in polynomial time and whose solution may still be verified in polynomial time. … Based on the definition alone, it is not obvious that NP-complete problems exist… The first natural problem proven to be NP-complete was the Boolean satisfiability problem, also known as SAT… However, after this problem was proved to be NP-complete, proof by reduction provided a simpler way to show that many other problems are also NP-complete, including the game Sudoku…a polynomial-time to Sudoku leads, by a series of mechanical transformations, to a polynomial-time solution of satisfiability, which in turn can be used to solve any other NP problem in polynomial time… In 1975, R. E. Ladner showed that if P≠NP, then there exist problems in NP that are neither in P or NP-complete. Such problems are called NP-intermediate problems. The graph isomorphism problem, the discrete logarithm problem, and the integer factorization problem are examples of problems believed to be NP- intermediate. …P means "easy" and "not in P" means "hard", an assumption known as Cobham's thesis. It is a common and reasonably accurate assumption in complexity theory; …There are algorithms for many NP-complete problems, such as the knapsack problem, the traveling salesman problem, Boolean satisfiability problem that can solve to optimality many real-world instances in reasonable time…Decades of searching have not yielded a fast solution to any of these problems, so most scientists suspect that none of these problems can be solved quickly. This, however, has never been proven (Wikipedia, 2021)".

III. Methodology

a) Definition of the Task

Any NP class problem can be solved by exhaustive search of all instances, i.e., by brute force search that requires exponential execution time, this is unacceptable in practice, therefore, we need to solve the NP problems in polynomial time, and if one of these NP problems is solved in polynomial time, then the others will also be solved in polynomial time. To solve the task where the worst-case run-time on an input of size n is $O(n^n)$ that have the highest growth rate, i.e., is greater than exponential and factorial time complexities that take $O(2^n)$ and O(!), we need to transform this task from infinitely exponential complexity class to polynomial complexity class using logarithmic looping of n^n if the value of n^n is explicit, but even then, when this task is solved, it will have no practical use as it leads to infinity only. Therefore, we need to solve the task of exponential time complexity that takes $O(2^n)$ to get the P vs. NP problem solution.

Exponential runtime complexity $O(2^n)$ is often seen in recursive functions that make 2 recursive calls that mean that growth doubles with each addition to the input data set (every problem in NP is recursive, and every recursive problem is recursively enumerable). Let us take, for example, a set with n elements, where we need to find (generate) all subsets of this set (the set theory is commonly used as a foundational system for the whole of mathematics and has various applications in computer science; its implications for the concept of infinity and its multiple applications have made set

theory a field of major interest; current research into the set theory covers a vast array of topics, ranging from the real number line structure to the study of the consistency; many mathematical concepts can be defined precisely using only set theory concepts). There are three ways to find the number of subsets of a set $S=\{a_0, a_1, \ldots, a_{n-1}\}$. The Tower of Hanoi is $O(2^n)$, as the expression of the trend we see would be $2^0+2^1+2^2+2^3+2^{n-1}=2^n-1$ that takes exponential running time. The second way is to translate between the binary representation of the rank and the subset when 1 means the corresponding element is in the subset, and 0 means the element is not in the subset, see below:

$$\text{We take } S=\{a_0, a_1, a_2\}.$$

Subsets of a given set:

0 000 { } the empty subset
1 001 $\{a_0\}$
2 010 $\{a_1\}$
3 011 $\{a_0, a_1\}$
4 100 $\{a_2\}$
5 101 $\{a_0, a_2\}$
6 110 $\{a_1, a_2\}$
7 111 $\{a_0, a_1, a_2\}$

As a result, we get exponential running time that will rising meteorically if we will add n elements to this set. And the third way, let us consider a set with 5 elements:

$$S=\{1, 2, 3, 4, 5\}.$$

What is the number of all possible and proper subsets of a given set with these 5 elements? There are 2^n subsets and 2^n-1 proper subsets that means that the number of all subsets of a set is 2^5 and the number of proper subsets is 2^5-1. To determine the Big-O runtime complexity, we do not need to look at how many recursive calls are made (iterating over all possible subsets of a set) since we will not deal with Fibonacci trees, it will be used only the task of the recursive Fibonacci number calculation that is $O(2^n)$, as the certain patterns in the recurrence relation lead to exponential results too (exponential time grows much faster than polynomial time). Therefore, we will get this using a new time complexity that works without a return (we capture one of the NP tasks in a sequence of matrix loops that runs in polynomial time and hack its secret arrangement without recursion). You need to read the paper at https://doi.org/10.3844/jcssp.2020.1610.1624 that is published recently and gets O(log n) complexity instead of $O(n^2)$ before continuing this reading since we will use this O(log n) complexity to solve this exponential task in polynomial running time (read this paper instead of the Methodology section, you can start reading at once from the end to clarify faster how it works, more exactly, see Lemma 21 and then other lemmas).

Let's continue if you have read. We will solve this NP problem using the new matrix model of computation concept and prove that this is a perfect path for its solution.

Lemma 1.0. The use of the NP task 2^n partitioning into 2^2 particles is a key for this NP task solution.

Proof. We need to generate a set of matrices to find the number of all subsets of this set with five elements that is $S=\{1, 2, 3, 4, 5\}$ (see above), where are 2^5 subsets that are equal to $2^2 \cdot 2^2 \cdot 2^1$. Each of these 2^2 particles gives one complete matrix. The matrices look like these matrix loops:

$$\text{Loop1,2,3}=\begin{pmatrix}2 & 2\\ 8 & 8\end{pmatrix}, \begin{pmatrix}2 & 2\\ 8 & 8\end{pmatrix}, \begin{pmatrix}2\\ 8\end{pmatrix},$$

where we have inserted these previous 2^2 particles of our partitioned task that is $2^5=2^2 \cdot 2^2 \cdot 2^1$ in an array as one of the options of this array to find the number of all possible subsets of a given above set, and the set of these matrices represents this decomposition of $2^5=2^2 \cdot 2^2 \cdot 2^1$, and each of them works to find these 2^2 particles, note that matrix Loop1,2,3 is not complete since the number of elements of the given set is odd, therefore, Loop 1,2,3 not works completely and carried over this incomplete matrix that is $\begin{pmatrix}2\\ 8\end{pmatrix}$ to the following loops; we are moving ahead only (without using backtracking to find all subsets), i.e., we do not need to iterate recursively (return), we take the result obtained by the first matrix loop and drag it to another matrix loop till we get to finish (terminate), and as we move ahead through the matrix loops, we cut the work at least in half and are closer to finding the last result, that is how we proceed, and further, we receive this:

$$\text{Loop2,3}=\begin{pmatrix}4 & 4\\ 6 & 6\end{pmatrix}, \begin{pmatrix}2\\ 8\end{pmatrix},$$

and finally, we have

$$\text{Loop3,0}=\begin{pmatrix}16 & 2\\ 84 & 98\end{pmatrix},$$

where the total $a=LE=2^5=(2\cdot2)\cdot(2\cdot2)\cdot2=(4\cdot4)\cdot2=16\cdot2=32=a_3$.

Remark 1.0. There is a reference map of these matrices that is:

$$\begin{pmatrix}L & E\\ I & T\end{pmatrix},$$

In the first loop, $L=E=2$, $I=T=sbasis-L=sbasis-E=8$, $sbasis=10$, $LE=4=a_1$, in the second loop, $L=E=4=a_1$, $I=T=sbasis-L=sbasis-E=6$,

sbasis=10, LE=16=a_2, in the third loop, L=16=a_2, E=2, I=sbasis-L=84, T=sbasis-E=92, sbasis=100, LE=32=a_3. We are interested only in the (a) options values, as all these elements of a given set are inserted on a position of (a)=LE options in this array after partitioning them into 2^2 equal particles, therefore, it is not necessary to determine the values of T elements, they can be dropped since these values will not be used in the main algorithm below. We use this algorithm for these 2^2 particle's logarithmic looping:

$$T(n)=(E:2\text{-}(E:2:(sbasis:(I\text{-}L))))\cdot sbasis$$

that provides the following

- Loop1,2,3 gives a_1=20:2-(20:2:(100:(80-20)))=4=2^2 - for Loop1,2 that goes to (4·4)·2
- Loop2,3 gives a_2=40:2-(40:2:(100:(60-40)))=16=4^2 - for Loop2 that goes to 16·2
- Loop3,0 gives a_3=200:2-(200:2:(10000:(8400-1600)))=32=16·2 - for Loop3

That means that the number of all subsets of a set is 32, including the empty subset, and the number of proper subsets is 32-1=31.

Remark 2.0. Keep in mind that we not only do not return to the matrix loop, where we already have received the result, we find the value of (a)=LE only for one complete matrix of each matrix loop since all complete matrices of each matrix loop are the same (they are copies).

Remark 3.0. Compare these steps with the following:

S={1, 2, 3, 4, 5}.

Subsets of a given above set:

{ }, {1}, {2}, {3}, {4}, {5}, {12}, {13}, {14}, {15}, {23}, {24}, {25}, {34}, {35}, {45}, {123}, {124}, {125}, {134}, {135}, {145}, {234}, {235}, {245}, {345}, {1234}, {1235}, {1245}, {1345}, {2345}, {{12345}.

Imagine how many returns (repeating moves) you will need to make to find all subsets of a set when the number of elements in a set is 20, 30, 80, etc.

Let's go further.

Lemma 2.0. The partitioning of the NP task 2^n into 2^2 particles remains a key for this NP task solution when the exponent of 2^n grows.

Proof. Suppose we need to find all subsets of a set with eight elements that is S={1, 2, 3, 4, 5, 6, 7, 8}, where are 2^8 subsets. The number of elements in this set is even, therefore, all matrices of the matrix Loop1,2.3.4 are complete. Further we have:

$$Loop1,2,3,4 = \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix},$$

then

$$Loop3,4 = \begin{pmatrix} 4 & 4 \\ 6 & 6 \end{pmatrix}, \begin{pmatrix} 4 & 4 \\ 6 & 6 \end{pmatrix},$$

and finally,

$$Loop4,0 = \begin{pmatrix} 16 & 16 \\ 84 & 84 \end{pmatrix}.$$

That is, there are 256 subsets in a given set.

Remark 1.1. Let us take a look at a visual model of this task that gives the scheme below. We have the following:

(2·2) (2·2) (2·2) (2·2) — Loop1,2,3,4, where L=E=2, I=8=sbasis-2, a_1=2·2=4

(4 · 4) (4 · 4) — Loop3,4, where L=E=4, I=6=sbasis-4, $a_2=a_1^2$=4·4=16

(16 · 16) — Loop4,0, where L=E=16, I=84=sbasis-16, $a_3=a_2^2$=16·16=256

256

The number of all subsets of a set that is 2^8 is equal to 256.

Theorem 1.0.

Regardless of how large the exponent of 2^n is, a sequence of matrix loops runs in polynomial time solving this exponential-time task, that means that an upper bound on the worst-case running time of this 2^n task is O(log n).

Corollary 1.0. We get a sequence of matrix loops that runs in polynomial time when we define the value of 2^n.

Proof

Let's go further and take a set with 30 elements, where we need to find all possible subsets of this set. The number of all subsets of this set is 2^{30}, and we have the following:

$$Loop1, \dots, 15 = \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}^{15\ complete},$$

then

$$Loop7{,}5, \dots, 15 = \begin{pmatrix} 4 & 4 \\ 6 & 6 \end{pmatrix}^{7\ complete}, \begin{pmatrix} 4 \\ 6 \end{pmatrix},$$

and then

$$Loop3{,}75, \dots, 15 = \begin{pmatrix} 16 & 16 \\ 84 & 84 \end{pmatrix}^{3\ complete}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix},$$

further,

$$Loop1{,}875, \dots, 15 = \begin{pmatrix} 256 & 256 \\ 744 & 744 \end{pmatrix}^{1\ complete}, \begin{pmatrix} 256 \\ 744 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \end{pmatrix},$$

and,

$$Loop1{,}875, \dots, 15 = \begin{pmatrix} 65536 & 256 \\ 34464 & dropped \end{pmatrix}, \begin{pmatrix} 16 & 4 \\ 84 & 96 \end{pmatrix},$$

and, finally,

$$Loop15, 0 = \begin{pmatrix} 16777216 & 64 \\ 83222784 & dropped \end{pmatrix}.$$

That is, there are 1 073 741 824 subsets in this set with 30 elements. We use this algorithm below for matrices of each matrix loop:

$$a{=}LE{=}(E{:}2{-}(E{:}2{:}(sbasis{:}(I{-}L)))){\cdot}sbasis$$

that takes O(log n) complexity.

Corollary 2.0. A sequence of matrix loops runs in polynomial running time when the exponent of 2^n increases and becomes larger.

Proof

As we need to estimate the asymptotic complexity of this 2^n task, let us consider, for instance, a set with 89 elements, where the number of all possible subsets is 2^{89}. We need to partition the 2^{89} into 2^2 particles, where are 44 complete and 1 incomplete matrices in the initial matrix loop (note that all incomplete matrices are carried to the following matrix loops until there are no complete matrices, then they are sequentially enclosed in additional matrix loops on the position of the (a) options in matrices), and we have the following:

$$Loop1, \dots, 44{,}5 = \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}^{44\ complete}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

then

$$Loop22{,}25, \dots, 44{,}5 = \begin{pmatrix} 4 & 4 \\ 6 & 6 \end{pmatrix}^{22\ complete}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

further,

$$Loop11{,}125,\ldots,44{,}5 = \begin{pmatrix} 16 & 16 \\ 84 & 84 \end{pmatrix}^{11\ complete}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and

$$Loop5{,}5625,\ldots,44{,}5 = \begin{pmatrix} 256 & 256 \\ 744 & 744 \end{pmatrix}^{5\ complete}, \begin{pmatrix} 256 \\ 744 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

we have

$$Loop2{,}78125,\ldots,44{,}5 = \begin{pmatrix} 65536 & 65536 \\ 34464 & 34464 \end{pmatrix}^{2\ complete}, \begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 256 \\ 744 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and then,

$$Loop1{,}390625,\ldots,44{,}5 = \begin{pmatrix} 4294967296 & 4294967296 \\ 5705032704 & 5705032704 \end{pmatrix}^{1\ complete}, \begin{pmatrix} 16777216 & 2 \\ 83222784 & 8 \end{pmatrix},$$

and finally,

$$Loop44{,}5{,}0 = \begin{pmatrix} 18446744073709551616 & 33554432 \\ 81553255926290448384 & dropped \end{pmatrix}.$$

This matrix looping gives a sequence that runs in polynomial time.

Loop1, …,44,5 gives the value of the (a_1) option for all complete matrices of this matrix loop, that is (2 · 2), and goes to the following matrix loop as the value of (L· E)=(4 · 4), and defines the number of all complete and incomplete matrices for the following matrix loop that is the Loop22,25, …,44,5 this number is (44,5:2)=22,25.

- a_1=(2:2-(2:2:(10:(8-2)))·10=4

Loop22,25, …,44,5 gives the value of the (a_2) option for all complete matrices of this matrix loop, that is (4 · 4), and goes to the following matrix loop as the value of (L · E)=(16 · 16), and defines the number of all complete and incomplete matrices for the following matrix loop that is the Loop11,125, …,44,5, i.e., (22,25:2)=11,125.

- a_2=(4:2-(4:2:(10:(6-4)))·10=16

Loop11,125, …,44,5 gives the value of the (a_3) option for all complete matrices of this matrix loop, that is (16 · 16), and goes to the following matrix loop as the value of (L· E)=(256 · 256), and defines the number of all complete and incomplete matrices of the Loop5,5625, …, 44,5, i.e., (11,125:2)=5,5625, etc.

- a_3=(16:2-(16:2:(100:(84-16)))·100=256
- a_4=(256:2-(256:2:(1000:(744-256)))·1000=65536
- a_5=(65536:2-(65536:2:(100000:(34464-65536)))·100000=4294967296
- a_6=(4294967296:2-(4294967296:2:(10000000000:(5705032704-4294967296)))·sbasis=2147483648-(2147483648:(10000000000:1410065408)))·10000000000=18446744073709551616
- a_7=(33554432:2-(33554432:2:(100000000000000000000:(81553255926290448384-18446744073709551616)))·sbasis=(16777216-(16777216:(100000000000000000000:631065118525808967 68)))·100000000000000000000=618970019642690137449562112=2^{89}

The number of all subsets of this set with 89 elements is 618970019642690137449562112.

Corollary 3.0. The sequence of matrix loops runs in polynomial time when the exponent of 2^n becomes extremely large.

Proof

Let us consider a set where the number of elements of this set is much larger than in previous sets. We take the set with 4117 elements, the number of all subsets of this set is 2^{4117} , and we have the following sequence of matrix loops:

$$Loop1,\dots,2058,5 = \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}^{2058\ complete}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

where we get the value of $a_1=(L \cdot E)= 2 \cdot 2=4$, (see the reference map of these matrices above), then,

$$Loop1029,25,\dots,2058,5 = \begin{pmatrix} 4 & 4 \\ 6 & 6 \end{pmatrix}^{1029\ complete}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

where we get the value of $a_2=a_1 \cdot a_1=16$, further,

$$Loop514,625,\dots,2058,5 = \begin{pmatrix} 16 & 16 \\ 84 & 84 \end{pmatrix}^{514\ complete}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and this matrix loop gives the value of $a_3=a_2 \cdot a_2=256$, and then,

$$Lopp257,3125,\dots,2058,5 = \begin{pmatrix} 256 & 256 \\ 744 & 744 \end{pmatrix}^{257\ complete}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

then we get the value of $a_4=a_3 \cdot a_3=65536$, further,

$$Loop128,65625,\dots,2058,5 = \begin{pmatrix} 65536 & 65536 \\ 34464 & 34464 \end{pmatrix}^{128\ complete}, \begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and we have the value of $a_5=a_4 \cdot a_4=4294967296$, and further,

$$Loop64,328125,\dots,2058,5 = \begin{pmatrix} 4294967296 & 4294967296 \\ 5705032704 & 5705032704 \end{pmatrix}^{64\ complete}, \begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and we get the value of $a_6=a_5 \cdot a_5=18446744073709551616$, then,

$$Loop32,1640625,\dots,2058,5 =$$

$$= \begin{pmatrix} 18446744073709551616 & 18446744073709551616 \\ 81553255926290448384 & 81553255926290448384 \end{pmatrix}^{32\ complete}, \begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

and we have the value of $a_7=a_6 \cdot a_6=340282366920938463463374607431768211456$, further,

$$Loop16,08203125,\dots,2058,5 =$$

$$= \begin{pmatrix} 340282366920938463463374607431768211456 & 340282366920938463463374607431768211456 \\ 659717633079061536536625392568231788544 & 659717633079061536536625392568231788544 \end{pmatrix}^{16\ complete},$$

$$\begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

then we get $a_8=a_7 \cdot a_7=115792089237316195423570985008687907853269984665640564039457584007913129639936$, and,

$$Loop8,041015625,\dots,2058,5 =$$

$$= \begin{pmatrix} 115792089237316195423570985008687907853269984665640564039457584007913129639936 & 1\dots \\ 884207910762683804576429014991312092146730015334359435960542415992086870360064 & 8\dots \end{pmatrix}^{8\ complete}, \begin{pmatrix} 65536 \\ 34464 \end{pmatrix}, \begin{pmatrix} 16 \\ 84 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

where we have $a_9=a_8 \cdot a_8$=1340780792994259709957402499820584612747936582059239337772356144372176403007354697680187429816690342769003185818648605085375388281194656994643364900608409 6 ..., note that the value of L element is always equal to E element and the value of I element is always equal to T element if these elements are the elements of one of the complete matrices of each matrix loop, further,

$$Loop4{,}0205078125,\ldots,2058{,}5 = \left(\begin{matrix}1340780792994259709957402499820584612747936582059239337772356144372176403007354697680187429816690342769003185818648605085375388281194656994643364900608409 6 \ldots \\ 8659219207005740290042597500179415387252063417940760662227643855627823596992645302319812570183309657230996814181351394914624611718805343005356635099391590 4 8 \ldots\end{matrix}\right)^{4\ complete},$$

$$\binom{65536}{34464},\binom{16}{84},\binom{2}{8},$$

and further, we get $a_{10}=a_9 \cdot a_9$=17976931348623159077293 0…224137216, the value of a_{10} is long enough to write it in full like the following values of a, then,

$$Loop2{,}0802534906425,\ldots,2058{,}5 = \left(\begin{matrix}1797693134862315907729330224137216 & 1797693134862315907729330224137216 \\ 1797693134862315907729330224137216 & 1797693134862315907729330224137216\end{matrix}\right)^{1\ complete},\binom{65536}{34464},\binom{16}{84},\binom{2}{8},$$

and then we have $a_{11}=a_{10} \cdot a_{10}$, and,

$$Loop1{,}005126953125 = \left(\begin{matrix}a_{11} & a_{11} \\ sbasis-a_{11} & sbasis-a_{11}\end{matrix}\right)^{1\ complete},\binom{65536}{34464},\binom{16}{84},\binom{2}{8},$$

and finally, we get the value of $a_{12}=a_{11} \cdot a_{11}$, where L=$a_{11}$, E=L=$a_{11}$, I=sbasis-L=sbasis-$a_{11}$, T=I=sbasis-$a_{11}$, then,

$$Loop2058{,}5{,}0 = \left(\begin{matrix}a_{12}=a_{11} \cdot a_{11}=(a_{11})^2 & (65536 \cdot 16 \cdot 2) \\ sbasis-a_{12} & dropped\end{matrix}\right).$$

We have the value of a_{total} that is equal to the value of ($a_{12} \cdot 65536 \cdot 16 \cdot 2$) = $2^{4\,111\,7}$, where 65536, 16 and 2 are the values from those incomplete matrices that were carried over all loops and the value of T element is dropped, we use this algorithm a=(E:2- (E:2:(sbasis:(I-L)))) to find these long values of all (a) options, and thus, regardless of how large the exponent of 2^n is, a sequence of such matrix loops runs in polynomial time.

Asymptotic analysis of the runtime of an algorithm that we use to find the value of (a) option for each complete matrix of these matrix loops is presented below.

Run-time analysis: Prove that (E:2-(E:2:(sbasis:(I-L))))·sbasis= O(log n). Let T(n) be the execution time for the input of size n, where $\lim_{n\to\infty}$, there exist positive constants and lower order terms that are not considered and can be omitted, then:

- T(n)=(E:2-(E:2:(sbasis:(I-L))))sbasis
- $T(n)=T_1(n)+T_2(n)+T_3(n)+T_4(n)+T_5(n)+T_6(n)=f(n)$
- $T_1(n)$=sbasis-L=I⇒O(n)
- $T_2(n)$=I-L⇒O(n)
- $T_3(n)$=sbasis:(I-L)⇒O(n)
- $T_4(n)$=E:2:(sbasis:(I-L))⇒O(n)
- $T_5(n)$=E:2-(E:2:(sbasis:(I-L)))⇒O(n)
- $T_6(n)$=(E:2-(E:2:(sbasis:(I-L))))·sbasis⇒can be omitted (dropped)

Let *f* and *g* be functions from positive numbers to positive numbers, where *f*(n)=(E:2-(E:2:(sbasis:(I-L))))·sbasis=O(n) and g(n)=O(log n). Prove the claim that *f*(n) is O(g(n)) if there exist positive constants c>0 and n_o>0 such that:

$$f(n) \le c*g(n) \text{ for all } n \ge n_0.$$

To prove big-O, we choose values for c and n_o and prove n>1 implies f(n)≤c*g(n):

1. Choose n_o=1,
2. Assuming n>1, find/derive a c such that:

$$\frac{f(n)}{g(n)} \le \frac{cg(n)}{g(n)} = c$$

that proves that n>1 implies f(n)≤c*gn. This means that function f(n) does not grow faster than g(n), or that function g(n) is an upper bound for f(n) for all sufficiently large $n \to \infty$.

An algorithm asymptotic running time is O(log n).

Notice. The value of sbasis is always equal to 10^n, therefore, we consider this value as an easy constant factor, and the I element is the 10's complement of the L element, therefore, it runs very quickly when we define the value of (sbasis-L).

Comparing the asymptotic running time:

An algorithm that runs in O(n) time is better than one that runs in O(2^n), and O(log n) is better than O(n).

Theorem 2.0.

It is enough to decompose n^n into the set of n^2 particles (fractions) to find the value of any n^n since there is an easy algorithm that solves the exponential-time task as the task that runs in polynomial time, i.e., we will turn (transform) NP to P using O{log n) complexity that will provide an easy solution for every n^2 particle of this set.

Proof

As any n^n can be easily decomposed (partitioned) into $n^2 \cdot n^2 \cdot \ldots \cdot n^2 \cdot n^1$ if the exponent of n^n is an odd number, and into $n^2\ n^2 \cdot \ldots\ n^2 \cdot n^2$ if the exponent of n^n is an even number, that we can find in logarithmic time using O(log n) complexity, hence we can find n^n in polynomial time, that means that P=NP. For example, it

is obvious, as we are aware, that $3^5=3^2\cdot3^2\cdot3^1$, or $5^6=5^2\cdot5^2\cdot5^2$, etc. The constant factors of this new algorithm will remain sustainable (steady) and scalable when n^n grows and goes to infinity. Let us consider the following:

Given a set (an array) of positive integers in matrix form:

$$\begin{pmatrix} n_1^n & n_2^n \\ n_3^n & n_4^n \end{pmatrix},$$

and let, for instance, $n_1=3$ and $n_2=2$, then $n_3=sbasis-n_1=7$ and $n_4=sbasis-n_2=8$, the values of the n_3 and n_4 elements of an array are the complements of the n_1 and n_2 elements, the sbasis=10, suppose we need to find the exponential values of these elements, when $n_1^n= 3^5$ and $n_2^n= 7^5$, the bases and the exponents of the n^n elements are taken arbitrarily, and the current sbasis is successive, then:

$$\begin{pmatrix} 3^5 & 2^5 \\ 7^5 & 8^5 \end{pmatrix},$$

and further, we decompose this array into this matrix of n^2 particles:

$$\begin{pmatrix} 3^2\cdot3^2\cdot3^1 & 2^2\cdot2^2\cdot2^1 \\ 7^2\cdot7^2\cdot7^1 & 8^2\cdot8^2\cdot8^1 \end{pmatrix}.$$

We got the decomposition of this n^n task into n^2 particles that transforms the exponential time to a polynomial that uses the new O(log n) complexity for $O(n^2)$, i.e., for these n^2 particles solving. Further, we make matrix loops for each of these n^2 particles that look like this:

$$\begin{pmatrix} 3 & 3 \\ 7 & 7 \end{pmatrix}\cdot\begin{pmatrix} 3 & 3 \\ 7 & 7 \end{pmatrix}\cdot\begin{pmatrix} 3 \\ 7 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}\cdot\begin{pmatrix} 2 & 2 \\ 8 & 8 \end{pmatrix}\cdot\begin{pmatrix} 2 \\ 8 \end{pmatrix}.$$

To solve these matrix loops, we will also use this algorithm:

$$a=n_1^n\cdot n_2^n=(n_2^n:2-(n_2^n:2:(sbasis:(n_3^n-n_1^n))))\cdot sbasis,$$

The reference matrix for n^n looks like this:

$$\begin{pmatrix} {n_1}^2\cdot{n_1}^2\cdot\ldots\cdot{n_1}^2\cdot n_1 & {n_2}^2\cdot{n_2}^2\cdot\ldots\cdot{n_2}^2\cdot n_2 \\ {n_3}^2\cdot{n_3}^2\cdot\ldots\cdot{n_3}^2\cdot n_3 & {n_4}^2\cdot{n_4}^2\cdot\ldots\cdot{n_4}^2\cdot n_4 \end{pmatrix}$$

for n^n, where the exponent of this n^n is an odd, and

$$\begin{pmatrix} {n_1}^2\cdot{n_1}^2\cdot\ldots\cdot{n_1}^2\cdot{n_1}^2 & {n_2}^2\cdot{n_2}^2\cdot\ldots\cdot{n_2}^2\cdot{n_2}^2 \\ {n_3}^2\cdot{n_3}^2\cdot\ldots\cdot{n_3}^2\cdot{n_3}^2 & {n_4}^2\cdot{n_4}^2\cdot\ldots\cdot{n_4}^2\cdot{n_4}^2 \end{pmatrix}$$

for n^n, where the exponent of this n^n is an even. These all are easy to check using any random instance of n^n.

IV. Results and Discussion

The major result of this paper is that $O(2^n)=O(\log n)$, that means that P=NP. Easy to solve (to find), easy to check (to verify), don't you think? This is a study that changes our understanding of a topic. We had to go beyond the rules for this. And it is easier than you think. There is no decision problem (a yes-no question) for the NP problem anymore, and we do not need the certificate for this, since we have simplified and eliminated all that was complicated multiple times over by various wrong theories and their numerous modifications, we no longer even need the SAT. The NP tasks do not require making two recursive calls when growth doubles with each addition to the input data set, we have a new path to solve this problem in polynomial running time using a sequence of matrix loops that uses a sorted array and takes O(log n) complexity. We get the desired reduction of the exponential problem to the polynomial problem. There are some known definitions of the P vs. NP problem that will be read in a new way in the future: 'The P versus NP problem is to determine whether every language accepted by some nondeterministic algorithm in polynomial time is also accepted by some (deterministic) algorithm in polynomial time (Cook, 2000)." 'P versus NP – a gift to mathematics from computer science (Smale, 2000)." "It is interesting to recall that the motivation for the development of the theory of computation, on which theoretical computer science is based, came from purely mathematical considerations. The paradox is emerging from Cantor's set theory emphasized the need to clarify the foundations of mathematics and, under Hilbert's leadership, concentrated attention on axiomatic proof systems. The quest to understand the power and limitations of axiomatizable systems led directly to the questions about all possible formal mechanical ways of deriving proofs (sequences with desired properties). In modern terms, it led to the search for what is and is not effectively computable (Hartmanis, 1989)." "The hope that mathematical methods employed in the investigation of formal logic would lead to purely computational methods for obtaining mathematical theorems goes back to Leibniz… (Davis & Putnam,1959)." "Your definition of experiments by using point-sets is perfectly satisfactory to me, I thought, however, that it might be good to say explicitly that a computation may be part of an "observation" (Neumann, 2005)." "The most comprehensive formal systems yet set up are, on the one hand, … and, on the other, the axiom system for set theory…These two systems are so extensive that all methods of proof used in mathematics today have been formalized in them (Gödel, 1931)." "Occasionally it happens that we seek the solution under insufficient hypotheses or in an incorrectly sense (Hilbert, 2000)." "The principal technique used for demonstrating that two problems are

related is that of "reducing" one to the other, by giving a constructive transformation that maps any instance of the first problem into an equivalent instance of the second (Garey, 1979)." "…any recognition problem solved by a polynomial time-bounded nondeterministic Turing machine can be reduced to the problem of determining whether a given proposition formula is a tautology (Cook, 1971)." "The class of languages recognizable by string recognition algorithms which operate in polynomial time is also invariant under a wide range of changes in the class of algorithms (Karp, 1972)." 'Due to the fact that no NP-complete problem can be solved in polynomial time… (Crescenzi & Kann, 1994)." "I offer a personal perspective on what it's about, why it's reasonable to conjecture that P≠NP is both true and provable… (Aaronson, 2011)." "…we can avoid brute - force search in many problems and obtain polynomial-time solutions. However, attempts to avoid brute force in certain other problems, including many interesting and useful ones, haven't been successful, and polynomial-time algorithms that solve them aren't known to exist (Sipser, 2012)." "As we solve larger and more complex problems with greater computational power and cleverer algorithms, the problem we cannot tackle begin to stand out (Fortnow, 2009)." "In recent years, the reducibility of computation in real environments to the standard Turing model has been brought increasingly into question (Cooper, 2004)." "The subject my talk is perhaps most directly indicated by simply asking two questions: first, is it harder to multiply than to add? and second, why? I grant I have put first of these questions rather loosely; nevertheless, I think the answer ought to be: yes. It is the second, which asks for a justification of this answer which provides the challenge (Cobham, 1965)." "Most of the computational problems that arise in practice turn out to be complete for one of a handful of complexity classes, even under very restrictive notions of reducibility (Agrawal, Allender, Impagliazzo, Pitassi, & Rudich, 2001)." "At present, when faced with a seemingly hard problem in NP, we can only hope to prove that it is not in P assuming that NP is different from P. (Goldreich, 2008)." "…an algorithm is any well-designed computational procedure that takes some value, or set of values, as input and procedures some value, or set of values, as output. An algorithm is thus a sequence of computational steps that transform the input into the output (Cormen, Lieserson, & Rivast, 2009)." "It is well known that every set in P has small circuits. Adelman was recently proved the stronger result that every set accepted in polynomial time by a randomized Turing machine has small circuits (Lipton & Karp, 1980)." 'I see complexity as the intricate and exquisite interplay between computation (complexity classes) and applications (that is, problem) (Papadimitriou, 1994)." "We do not know of polynomial-time algorithms for these problems, and we cannot prove that polynomial-time algorithms exist…These are the NP-complete problems… (Kleinberg & Tardos, 2006)." "…there is a strictly ascending sequence with a minimal pair of upper bounds to the sequence…if P≠NP then there are members of NP-P that are not polynomial complete (Ladner, 1975)." "…minimal propositional logic corresponds to dependent simply typed-calculus… (Sorensen & Urzyczyn, 1998)." 'Practical problems requiring polynomial time are almost solvable in an amount of time that we can tolerate, while those that require exponential time generally cannot be solved except for small instances (Hopcroft, Motwani, & Ullman, 2001)." "Some success was had by causing the machine to systematically eliminate the redundancy; but the problem of total length increasing rapidly still remained when more complicated problems were attempted (Davis, Logemann, & Loveland, 1961)." "Gȍdel and others went on to show that various other mathematically interesting statements, besides the consistency statement, are undecidable by P, assuming it to be consistent… (Boolos, Burgess, & Jeffrey, 2007)." "There has been much work in getting the number of variables needed for an undecidability result to be small (Gasarch, 2021)."

V. Conclusion

It is possible to solve the exponential-time task in polynomial time if we forget recursion that takes $O(2^n)$ complexity and goes to regress only. As you see, it is clear that the new notion of the decision procedure for the NP problem exists. We have a completely new definition of the certificate for this NP problem that can not only be checked in polynomial time but also solved in polynomial time. And there is no case when this new uniform procedure is not valid, the algorithm terminates with a correct answer on any input instance of 2^n and does not involve seeking forever (without Halting problem, without approximation), i.e., this new certificate is consistent, therefore, we solve this NP task rapidly, accurately, and easily using this unthinkably easy computational tactic above.

Acknowledgment

I would like to express my special thanks of gratitude to the reviewers for their thoughtful comments and the effort and expertise that they contribute to review the manuscript.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The corresponding author has NO conflicts of interest to disclose.

Ethics

This article is original and contains unpublished material.

References Références Referencias

1. Aaronson, S. (2017). P= ?NP. (PDF)
2. Agrawal, M., Allender, E., Impagliazzo, R., Pitassi, T., & Rudich, S. (2001). Reducing the complexity of reductions. Birkhauser Verlag, Basil. Computational complexity, 10 (2001), 147-138. (PDF)
3. Arora, S., & Barak, B. (2009). Computational complexity: a modern approach. Cambridge University Press.
4. Boolos, G. S., Burgess J. P., & Jeffrey R. C. (2007). Computability and logic. Cambridge University Press.
5. Cobham, A. (1965). The intrinsic computational difficulty of functions. (PDF)
6. Cormen, I., Leiserson, C, Rivast, R., & Stein, C. (2009). Introduction to algorithms. The MIT Press. Cambridge, Massachusetts. London, England. (PDF)
7. Cook, S. A. (1971). The complexity of theorem-proving procedures. University of Toronto. (PDF)
8. Cook, S. A. (April 2000). "The P versus NP Problem". Clay Mathematics Institute. Retrieved 18 October 2006.
9. Cooper, S. B. (2004). Computability theory. University of Leeds, U. K. (PDF) Crescenzi, P., & Kann, V. (1994). A compendium of NP optimization problems. (PDF)
10. Chotchaeva, Z. A. (2020). The new matrix model of computation based purely on quite a new concept of the matrix computations for extremely quick web pages loading. Journal of Computer Science, 16(11), 1610-1624.
11. Davis, M., & Putnam, H. (1959). A computing Procedure for Quantification theory. Journal of the ACM, Volume 7.
12. Davis, M., Logemann, G., & Loveland, D. (1961). A Machine Program for theorem proving. Communications of the ACM, Volume 5, Issue 7.
13. Fortnow, L. (2009). The status of the P vs NP. Communications of the ACM, 52(9); 78-86. Garey, M. R. & Johnson, O. S. (1979). Computers and Intractability (Vol. 174). San Francisco: freeman.
14. Gasarch, W. (2002). Hilbert's Tenth Problem: Refinements and Variants. arXiv:2104.07220v2 [math.LO]
15. Gȍdel, K. (1931). On formally undecidable propositions of principia mathematics and related systems. (PDF)
16. Goldreich, O. (2010). P, NP, and NP-Completeness: The Basics of Computational Complexity. Cambridge University Press.
17. Hartmanis, J. (1989). Gödel, von Neumann, and the P=?NP problem. Bulletin of the European Association for Theoretical Computer Science. 38: 101-107.
18. Hilbert, D. (2000). Mathematical problems, AMS, Volume 37. (PDF)
19. Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2001). Introduction to automata theory, languages, and computations. Addison-Wesley. (PDF)
20. Karp, R.M. (1972). Reducibility among Combinatorial Problems. DOI: 10.1007/978-3-540-68279-0_8, Springer, 2010.
21. Kleinberg, J., & Tardos, E. (2005). Algorithm design. Cornell University. (PDF)
22. Ladner, R. E. (1975). On the structure of Polynomial-time reducibility. Journal of the ACM. Volume 23, Issue 1.
23. Lipton, R. J., & Karp. R. M. (1980). Some connections between nonuniform complexity classes. Proceedings of the ACM, 302-309.
24. Morten Heine B. Sorensen, & Pawel Urzyczyn (1998). Lectures on the Curry-Howard Isomorphism. University of Copenhagen & Warsaw.
25. Neumann, von J. (2005). Selected letters. History of mathematics, Volume 27. (PDF)
26. Papadimitriou, C. (1993). Computational Complexity. Addison-Wesley.
27. Sipser, M. (2012). Introduction to the Theory of Computation. Cengage learning.
28. Smale, S. (2000). Mathematical problems for the next century. The mathematical intelligencer 20 (2). DOI:10.1007/BF03025291
29. Wikipedia. (2021). P versus NP problem. Wikipedia.

This page is intentionally left blank

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 3 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Statistical Literacy among Practicing Clinicians from India: A Brief Survey

By Ankush Jindal, Dr. Manishi Bansal & Dr. Anupam Jindal

Abstract- *Objectives:* To understand statistical literacy among practicing clinicians as well as to take suggestions on its implementation in medical curriculum.

Methods: A web based online survey was conducted among practising clinicians of India. The questionnaire consisted of 30 questions in three parts. Part A included questions on general information (5 questions), Part B included questions on application of statistical concepts (20 questions), and Part C asked for opinions on integration of biostatistics in medical curriculum (5 questions). Part B questions were rated on a 5-point Likert scale in which 1 indicated no confidence and 5 indicated complete confidence.

Results: A total of 416 clinicians responded to the questionnaire. Complete confidence in the use of SPSS software was seen in 15.8% whereas 26.7% had no confidence in it. The highest confidence was seen in statistical equations like graphical representation of data (44.7%), and sensitivity and specificity (45.2%) whereas lowest confidence was seen in COX proportional hazard regression (12.9%) and ROC curves (11.7%). Out of 416 clinicians, 136 (32.4%) had done training in statistics at undergraduate level, 128 (30.5%) did self-learning and 152 (37.1%) had received no formal training.

Keywords: *statistical knowledge, medical curriculum, evidence-based medicine, critical appraisal, medical students.*

GJCST-G Classification: *I.4.10*

Strictly as per the compliance and regulations of:

Statistical Literacy among Practicing Clinicians from India: A Brief Survey

Ankush Jindal [α], Dr. Manishi Bansal [σ] & Dr. Anupam Jindal [ρ]

Abstract- Objectives: To understand statistical literacy among practicing clinicians as well as to take suggestions on its implementation in medical curriculum.

Methods: A web based online survey was conducted among practising clinicians of India. The questionnaire consisted of 30 questions in three parts. Part A included questions on general information (5 questions), Part B included questions on application of statistical concepts (20 questions), and Part C asked for opinions on integration of biostatistics in medical curriculum (5 questions). Part B questions were rated on a 5-point Likert scale in which 1 indicated no confidence and 5 indicated complete confidence.

Results: A total of 416 clinicians responded to the questionnaire. Complete confidence in the use of SPSS software was seen in 15.8% whereas 26.7% had no confidence in it. The highest confidence was seen in statistical equations like graphical representation of data (44.7%), and sensitivity and specificity (45.2%) whereas lowest confidence was seen in COX proportional hazard regression (12.9%) and ROC curves (11.7%). Out of 416 clinicians, 136 (32.4%) had done training in statistics at undergraduate level, 128 (30.5%) did self-learning and 152 (37.1%) had received no formal training. All the responders agreed that biostatistics should be included in medical curriculum. Seventy-five percent clinicians believed that under graduation is the apt time to learn medical statistics while 20.2% wanted to learn at postgraduate level.

Conclusions: This study found considerably low levels of statistical literacy among practising clinicians. However, they are keenly interested towards its integration in undergraduate medical curriculum.

Keywords: statistical knowledge, medical curriculum, evidence-based medicine, critical appraisal, medical students.

I. Introduction

As a fourth year medical student, I found difficulty in interpretating the results of research articles. I took assistance from my parents (who happen to be well known clinicians in their respective fields) but unfortunately they were also ignorant about the details of the statistics used. I searched our syllabus of medical education till final year and found no defined syllabus for medical statistics. I studied about the scenario of medical statistics in medical curriculum and its outcomes, and this led me to formulate this study.

Author α: Fourth year MBBS student, Government Medical college and hospital, Chandigarh, India.
Corresponding Author σ: Senior Consultant Radiation Oncology, Fortis Hospital, Mohali, India. e-mail- manishi1@yahoo.com
Author ρ: Additional Director Neurosurgery, Fortis Hospital, Mohali, India.

In the era of evidence-based medicine, it is very pertinent for clinicians to critically appraise the published literature in terms of design, conduct and analysis of the study so as to logically interpret the results (McColl, 1998 &Morris, 2002) This requires a fundamental knowledge of biostatistics which is lacking to a variable extent in practicing physicians as seen in several surveys conducted in 1980s (Weiss, 1980 &Wuff, 1987). The problem has become more apparent in recent times because of the use of complicated statistical method, which has interpretation of results in only 21% of the published articles (Horton et al., 2005). It has been already suggested by Palmer that 21st century doctors will need an armoury of critical appraisal skills to assess the research data (Palmer, 2002). Keeping this background in mind, we conducted a survey with the main objective to assess the knowledge of the basic methods of research and data analysis among medical doctors in India and to get suggestions from practicing doctors as to how and when statistics should be integrated to medical curriculum.

II. Materials and Methods

A web based online survey using Google web-application was conducted between October to December 2020 among practising clinicians of various fields in government and private sector in India. The survey was floated on social media (WhatsApp) among various groups and they were informed that the results of the survey might be used for analysis and medical publication. The participation was voluntary with no compulsion and was not limited to any institution or geographic area. The respondents' anonymity was ensured.

The questionnaire consisted of 30 questions in three parts. Part A included questions on general information and demographics (5 questions), Part B included questions on detailed knowledge and application of statistical concepts in medical research (20 questions), and Part C asked for opinions on integration of biostatistics in medical curriculum (5 questions). Fourteen questions in part B were rated on a 5-point Likert scale in which 1 indicated no confidence and 5 indicated complete confidence.

III. Statistical Analysis

The data collected was transferred to MS Excel data sheet. Data analysis was performed using SPSS (Statistical Package for the Social Sciences) version 22.0developed by IBM Corporation. Qualitative data was expressed using frequency and percentage. Quantitative data was explained using descriptive statistics. To compare the relation of different statistical concepts with variables, Chi-square test was used. P value of 0.05 was considered as statistically significant.

IV. Results

A total of 416 clinicians responded to the questionnaire over a period of three months. The results of different sections of the questionnaire are as follows-

Part A

Out of 416 clinicians, 272 (65.3%) were men and 144 (34.6%) were women with age varying from 25 years to 71 years. The mean age was 46.7 years and 224 (53.8%) clinicians were in the age group 45-55 years. Clinicians practicing oncology were 128 (30.7%), followed by paediatrics (7.69%), critical care and medicine (6.7% each) and rest were from gynaecology, neurosurgery, cardiology, and other clinical and non-clinical specialities [Table 1]. Most of the consultants were from private sector (57.1%) and 21.9% each from institutional and government sector. Years of practice ranged from 1 to 48 years with an average of 18.4 year and 14.4% had more than 20 years of practice in their respective fields.

Table 1: Clinical specialties for survey respondents

Clinical Specialty	Number	Percentage (%)
Oncology	128	30.7
Pediatrics	32	7.7
Anesthesia	28	6.7
Medicine	28	6.7
Gynecology	20	4.8
Neurosurgery	20	4.8
Cardiology	20	4.8
Orthopedics	20	4.8
Pathology	16	3.8
ENT	16	3.8
Ophthalmology	16	3.8
Nephrology	12	2.8
General surgery	12	2.8
Radiodiagnosis	8	1.9
Dermatology	8	1.9
Others	32	7.7

Part B

284 clinicians (68.3%) have done clinical research while 132 (31.7%) have never been involved in any clinical research so far. The number pf publications by the clinicians ranged from none to 184 in number with an average of 19.4 publications. When asked about the general understanding of all the statistical terms when reading a research article, only 10.3% were completely confident in their understanding whereas 4.8% were not at all confident (Table 2). The majority (42%) rated average confidence. However, 43.3% felt the relevance of biostatistics in medical curriculum (Likert scale 5). Only 17.3% clinicians (with complete confidence) indicated that they use statistical information in forming opinions or when taking decisions in medical care whereas 44.2% had more than average confidence on this question (Likert 4). Majority of the respondents (91.5%, Likert 4 and 5) agreed that to be an intelligent reader, it is necessary to know something about statistics.

Knowledge of fourteen statistical concepts was assessed on a Likert scale of 1 to 5 [table 2]. The results were as follows- understanding P value with complete confidence in 32.7% and more than average confidence in 31.7%. Confidence interval was completely understood in 25.2% and more than average in 33%. For standard deviation, 35.9% and 36.9% were completely confident and more than average confident respectively. Complete confidence in understanding of graphical presentation of data was seen in 44.7% of clinicians, survival analysis in 30.8% whereas it was only 11.7% for ROC curves(lowest respondents) and 13.3% for cluster analysis. Complete confidence in the use of software like SPSS was seen in 15.8% and near complete

confidence in 23.8% of the responders whereas 26.7% had no confidence at all in its use. Sensitivity and specificity in a data could be interpretated completely in highest number of respondents (45.2%), laws of probability in 21.4 % and summarizing and analysing missing data in 15.5%. Regression analysis was completely interpretated in only 15.8% whereas 20.8% had no knowledge about it. COX proportional hazard regression was seen with complete confidence in 12.9% and no confidence in 28.7% of the responders. More than average confidence in chi-square test (29.7%) and 9.9 % had no confidence at all. Most of the statistical concepts were rated as average confidence (Likert scale 3).

Table 2: Knowledge on statistical concepts among clinicians

Questions	Lowest confidence n (%)	Little more confidence n (%)	Average confidence n (%)	More than average confidence n (%)	Highest confidence n (%)
When reading a research article do you understand all the statistical terms mentioned (n = 416)	20 (4.8)	71 (17.0)	175 (42.0)	107 (25.7)	43 (10.3)
How do you perceive the relevance of biostatistics in medical curriculum (n = 416)	4 (1)	8 (1.9)	60 (14.4)	164 (39.4)	180 (43.3)
I often use statistical information for forming opinions or making a decision in medical care (n = 416)	12 (2.9)	40 (9.6)	108 (26)	184 (44.2)	72 (17.3)
To be an intelligent reader is it necessary to know statistics? (n = 416)	4 (1)	12 (2.9)	36 (8.7)	164 (39.4)	200 (48.1)
p- value (n = 416)	24 (5.8)	48 (11.1)	76 (18.3)	132 (31.7)	136 (32.7)
Confidence interval (n =412)	36 (8.7)	52 (12.6)	84 (20.4)	136 (33)	104 (25.2)
Standard deviation (n=412)	16 (3.9)	20 (4.9)	76 (18.4)	152 (36.9)	148 (35.9)
Graphical presentation of data (n=412)	12 (2.9)	8 (1.9)	32 (7.8)	176 (42.7)	184 (44.7)
Survival analysis (n=416)	40 (9.6)	32 (7.7)	80 (19.2)	136 (32.7)	128 (30.8)
ROC curve (n=412)	64 (15.5)	68 (16.5)	112 (27.2)	120 (29.1)	48 (11.7)
Cluster analysis (n=408)	68 (16.7)	80 (19.6)	92 (22.5)	112 (27.5)	56 (13.7)
Use of software like SPSS (n=404)	108 (26.7)	44 (10.9)	92 (22.8)	96 (23.8)	64 (15.8)
Sensitivity and Specificity (n=416)	12 (2.9)	28 (6.7)	56 (13.5)	132 (31.7)	188 (45.2)
Laws of probability (n=412)	28 (6.8)	52 (12.6)	108 (26.2)	136 (33)	88 (21.4)
Summarizing and analyzing missing data (n=412)	52 (12.6)	76 (18.4)	108 (26.2)	112 (27.2)	64 (15.5)
Regression analysis (n=404)	84 (20.8)	80 (19.8)	112 (27.7)	64 (15.8)	64 (15.8)
COX proportional hazard regression (n=404)	116 (28.7)	68 (16.8)	80 (19.8)	88 (21.8)	52 (12.9)
Chi-square test (n=404)	40 (9.9)	64 (15.8)	96 (23.8)	120 (29.7)	84 (20.8)

Part C

When asked about any previous training done in medical statistics, 136 (32.4%) responded that they did it as part of undergraduate curriculum, 128 (30.5%) did self-learning and 152 (37.1%) had received no formal training in statistics [Figure 1]. All the responders agreed that biostatistics should be included in medical curriculum and 92.3% were interested to learn more about it if given a chance. Seventy-five percent clinicians believed that MBBS is the apt time to learn medical statistics while 20.2% wanted to learn during junior residency [Figure 2]. A varied number of suggestions were given when asked about how to improve biostatistics training among doctors.

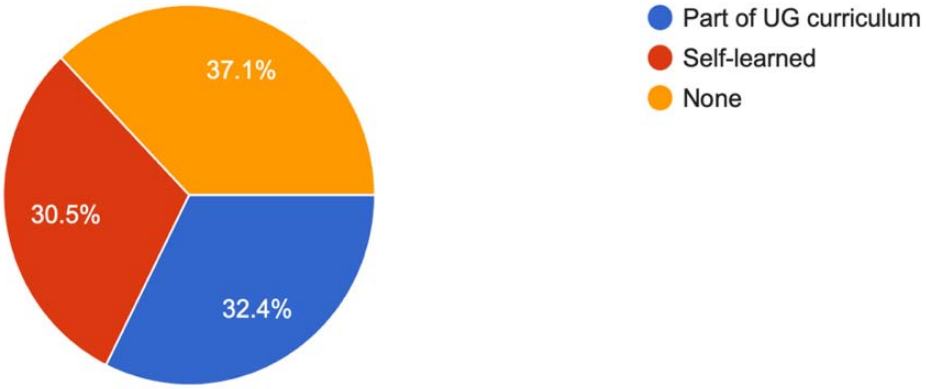


Figure 1: Any previous training/course attended on medical statistics?

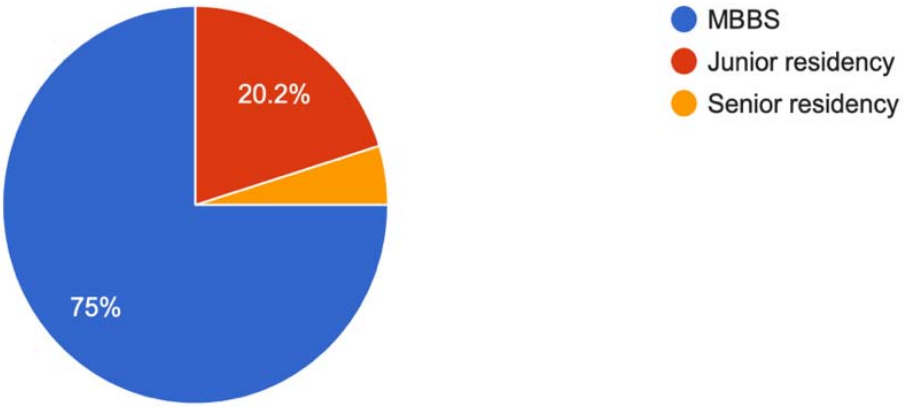


Figure 2: When should medical statistics be taught?

V. Discussion

Medical statistics (Biostatistics) has played an integral role in modern medicine. Statisticians help researchers design studies, analyse data from medical experiments, help interpret the results of the analyses, and collaborate in writing articles to describe the results of medical research (Google Scholar). However, statistics is full of concepts and technical terms which may be difficult to understand and this presents an important barrier to knowledge use. Also anecdotal experience supports that statistics is not the most liked subject in the undergraduate medical curriculum (Altman et al., 1991 & Freeman, 2008). To bridge this gap we need to integrate biostatistics in medical curriculum either at graduate or post graduate level (Editorial, Lancet, 2007).

We developed a basic survey questionnaire to assess the knowledge of statistics among practicing clinicians and reflect the statistical methods and results most represented in contemporary research studies. Our results suggest that only a limited number of clinicians were completely confident in using statistical equations and mostly scaled on average or below average on a Likert scale. This correlates well with the lesser confidence in the use of SPSS software (26.7% had lowest confidence and only 15.8% were completely confident). The highest confidence was seen in statistical equations like graphical representation of data (44.7%), and sensitivity and specificity (45.2%) which is the basic statistical concept whereas the lowest confidence was seen in COX proportional hazard regression (12.9%) and ROC curves (11.7%) which are relatively difficult concepts. The poor knowledge of statistical terms in our study reveals insufficient training of the clinicians in the past. Nearly 37.1% had received no formal training in statistics as opposed to a study by Windish et al where more than 68% of respondents had

some training in biostatistics (Windish et al., 2007). Another reason might be the lesser involvement of clinicians in research activities which was 31.7% in our study, this number is far more than a study by Susan et al where10 % had never been involved in any health research (Miles et al.,2010). This may be due to fact that our cohort is diverse with respect to age, clinical experience and type of practice (government or private).

We found that a better knowledge of biostatistics in clinicians was associated with their prior training in statistics (either part of medical curriculum or self-learning), more years of clinical experience and more number of publications, although no statistical correlation could be found. Similar finding was seen in astudy by Novack et al (Novack et al., 2006). Respondents with higher confidence in their statistical knowledge performed better on the questions pertaining to statistical concepts in part B, also seen in Windish study (Windish et al., 2007).

Our findings suggest that all the doctors recognised the value of undergraduate training in statistics and majority (92.3%) have the desire to learn even now, which indicates the relevance of the topic. Similar results were seen in a study by Windish in which 95% responders agreed that to be an intelligent reader it is necessary to know statistics (Windish et al., 2007). More than 58% responders in their study indicated that they use statistical information in forming opinions and in our study it was seen in 61.5% responders (Likert 4 and 5).

Seventy five of the participants believed that medical statistics should be incorporated into undergraduate medical curriculum. This is very well established in various studies worldwide that the foundation years are the best to introduce any new syllabus for better understanding of the subject (McColl, 1998 & Miles, 2010). The clinicians offered informative suggestions as to how undergraduate statistical training can be improved. First of all, medical statistics can be introduced along with epidemiology early in the undergraduate training. The main aim of the course is to understand the conceptual basis and usage of common statistical methods, and their application in clinical medicine (Swift et al., 2009). The teaching needs to ensure that medical students appreciate the relevance of learning a new skill. Secondly, it should be more interactive and practical oriented. The biostatistics course can be divided into small group tutorial based sessions based on one or more problems which contains both statistical and epidemiological data (Astin et al., 2002). The emphasis is on enabling students to critically appraise research and other evidence. In a study by Parkes, critical appraisal teaching resulted in a significant improvement in critical appraisal knowledge as compared to 6% improvement in control group (Parkes et al., 2002). Students can be provided with access to a computer assisted learning package which they can access freely. Finally the course can be concluded with a short examination on statistical methods. Later in the fourth year, some advanced topics such as meta-analysis and a project in which students can critically appraise research papers can be included in the curriculum. Furthermore, previously learned statistical concepts should be regularly reinforced throughout career with clinically integrated interactive teaching. This was seen in a survey Looney et al in which it was found that more than 90% of medical schools focussed their biostatical teaching in preclinical years without later reinforcement (Looney et al., 1998). On the contrary, few clinicians believed that medical students are already overburdened with their syllabus and there should not be extra subjects besides the existing ones. This implies that post graduation is better time for teaching other subjects as this time is ideal as they are more focussed on their clinical work and writing thesis. However interested students can learn it during vacation time or from online courses but first the students should be appraised of the need of such topic.

Our study has limitations, firstly the study cohort diversity. There is a diverse group of practising clinicians in terms of age, various specialties involved with different level of experience and type of practice. Secondly, our survey was purposely kept brief thus limiting our ability to assess understanding of all biostatistical concepts in detail. Nonetheless, our study is the first of its kind involving a large number of clinicians from India and it helps in providing useful information about the basic statistical knowledge among the practicing clinicians.

VI. Conclusions

The results of this study suggest that knowledge of statistical software and statistical concepts is lacking to various extent among practicing clinicians of India. However, they are keenly interested to learn more about it even at any stage of their career. There is more favour towards integration of statistical literacy in undergraduate curriculum so as to form a firm base in those years. It involves learning of new skills, almost a new language, and thus a more interactive form of teaching is necessary in which problems and methods can be discussed (Barleyy et al., 2016). Small group teaching sessions are therefore more appropriate for this. It is pertinent to not only make the teaching explicitly relevant to future practice but also implies the need for more robust training in biostatics among medical graduates.

Conflict of interest
None.

Financial support
Nil

Acknowledgement

The authors are thankful to the clinicians who participated in the study.

References Références Referencias

1. Altman, DG. & Bland JM. (1991). Improving doctors' understanding of statistics. J R Stat Soc Ser A, 154, pp. 223-67.
2. Astin, J. Jenkins, T. & Moore L. (2002). Medical students' perspective on the teaching of medical statistics in the undergraduate medical curriculum. Statistics in Medicine, 21(7), pp. 1003-1006.
3. Barlett, G. & Gagnon, MA. (2016). Physicians and knowledge translation of statistics: mind the gap. Canadian Medical Association Journal, 188(1), pp. 11-12.
4. Doctors and medical statistics. (2007). The Lancet, 370, pp. 910.
5. Freeman, JV., Collier, S.,Staniforth, D., &Smith,KJ. (2008). Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students. BMC Med Educ, 8, pp. 28.
6. Horton ,NJ & Switzer, SS.(2005). Statistical methods in the journal. N Eng J Med, 353(18), pp. 1977-1979.
7. Looney, SW., Grady, CS., & Steiner, RP. (1998). An update on biostatistics requirement in U.S. medical schools. Acad Med, 73(1), pp.92-94.
8. McColl, A., Smith, H., White, P., & Field, J. (1998). General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. BMJ, 316, pp. 361-365.
9. Miles, S., Price, GM., Swift, L., Shepstone, L., and Leinster., SJ. (2010). Statistics teaching in medical school: Opinions of practicing doctors. BMC Med Educ, 10, pp. 75.
10. Morris, RW., (2002). Does EBM offer the best opportunity yet for teaching medical statistics? Statistics in Medicine, 21(7),pp. 969-977.
11. Novack, L., Jotkowitz, A., Knyazer, B., Novack, V. (2006). Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors. Postgrad Med J, 82, pp. 817-822.
12. Palmer, CR. (2002). Discussion: teaching hypothesis tests: time for significant change? Statistics in Medicine, 21(7), pp. 995-999.
13. Parkes, J., Hyde, C., Deeks, J., et al. (2002). Teaching critical appraisal skills in health care settings. Cochrane Rev, 2. Oxford: England.
14. Swift, AL., Miles, S., Price, GM., Shepstone, L., Leinster, S. (2009). Do doctors need statistics? Doctors use of and attitudes to probability and statistics. Stat Med, 28, pp. 1969-81.
15. Weiss, ST & Samet, JM. (1980). An assessment of physician knowledge of epidemiology and biostatistics. J Med Educ, 55(8), pp. 692-697.
16. Windish, DM., Huot,SJ., Green, ML. (2007). Medicine residents' understanding of the biostatistics and results in the medical literature. Journal of the American Medical Association, 298(9), pp. 1010-1022.
17. Wulff, HR., Anderson, B., Brandenhoff, P., Guttler, F. (1987). What do doctors know about statistics? Stat Med, 6(1), pp. 3-10.

https://www.biostat.washington.edu/about/biostatistics (last accessed on 02/03/2021)

Global Journal of Computer Science and Technology: G Interdisciplinary
Volume 21 Issue 3 Version 1.0 Year 2021
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Proposed Method to Identify the Occurrence of Diabetes in Human Body using Machine Learning Technique

By Tanvir Rahman
Stamford University

Abstract- Advanced machine-learning techniques are often used for reasoning-based diagnosis and advanced prediction system within the healthcare industry. The methods and algorithms are based on the historical clinical data and factbased Medicare evaluation. Diabetes is a global problem. Each year people are developing diabetes and due to diabetes, a lot of people are going for organ amputation. According to the World Health Organization (WHO), there is a sharp rise in number of people developing diabetes. In 1980, it was estimated that 180 million people with diabetes worldwide. This number has risen from 108 million to 422 million in 2014. WHO also reported that 1.6 million deaths in 2016 due to diabetes. Diabetes occurs due to insufficient production of insulin from pancreas. Several research show that unhealthy diet, smoking, less exercise, Body Mass Index (BMI) are the primary cause of diabetes. This paper shows the use of machine learning that can identify a patient of being diabetic or non-diabetic based on previous clinical data. In this article, a method is shown to analyze and compare the relationship between different clinical parameters such as age, BMI, Diet-chart, systolic Blood Pressure etc. After evaluating all the factors this research work successfully combined all the related factors in a single mathematical equation which is very effective to analyze the risk percentage and risk evaluation based on given input parameters by the participants or users.

GJCST-G Classification: H.1.2

A Proposed Method to Identify the Occurrence of Diabetes in Human Body using Machine Learning Technique

Tanvir Rahman

Abstract- Advanced machine-learning techniques are often used for reasoning-based diagnosis and advanced prediction system within the healthcare industry. The methods and algorithms are based on the historical clinical data and fact-based Medicare evaluation. Diabetes is a global problem. Each year people are developing diabetes and due to diabetes, a lot of people are going for organ amputation. According to the World Health Organization (WHO), there is a sharp rise in number of people developing diabetes. In 1980, it was estimated that 180 million people with diabetes worldwide. This number has risen from 108 million to 422 million in 2014. WHO also reported that 1.6 million deaths in 2016 due to diabetes. Diabetes occurs due to insufficient production of insulin from pancreas. Several research show that unhealthy diet, smoking, less exercise, Body Mass Index (BMI) are the primary cause of diabetes. This paper shows the use of machine learning that can identify a patient of being diabetic or non-diabetic based on previous clinical data. In this article, a method is shown to analyze and compare the relationship between different clinical parameters such as age, BMI, Diet-chart, systolic Blood Pressure etc. After evaluating all the factors this research work successfully combined all the related factors in a single mathematical equation which is very effective to analyze the risk percentage and risk evaluation based on given input parameters by the participants or users.

I. Introduction

a) Background

Generally, Diabetes Mellitus (DM) develops in the body silently when there are higher or uncontrolled blood glucose level exists in the blood-plasma cell for a long time. Food is the prime source of calorie and food is prime the energy generator of the body. Generally, the foods are taken in regular basic contain a lot of glucose or glucose substance. Glucose is the primary and basic unit of energy circulation and energy regulation. Glucose is divided into several substances and then the small cell units are oxidized with the sufficient amount of oxygen. Then, the small subsequent oxygen particles are transmitted through blood circulation and produce sufficient amount of energy and nutrition for all the organs of the body. Insulin is a pancreas produced Hormone which is the key component to synthesis Glucose and divide Glucose into millions of active particles. For a healthy and active person sufficient amount of Insulin is produced and emitted from Pancreas. That is why for a general Non-diabetic patient , Insulin production rate is equal to the glucose Intake of the body .So, For a Non diabetic-patient the all the amount glucose casted from daily food intake ,is sufficiently divided into molecules and produces energy and rest of the unused energy is stored as Fat in the body. In the common scenario, no extra glucose particles are available in the blood plasma. According to several health study, a person is considered to be a Type-1 diabetic patient when his/her pancreas fails to generate sufficient amount of insulin to react with glucose.

The normal range of glucose level is reference value is 3.9 to 5.4 mmol/l (70 to 99 mg/dl) [1] for normal patients at fasting phase and according to American Diabetes Association the reference value at fasting time in the period of Diagnosis of diabetes is considered as 7.0 mmol/l (126 mg/dl) or above [1] and the reference value for non-diabetic patients is under 7.8 mmol/L and preferred value for Type-2 diabetes is under 8.5 mmol/L at Random diabetes testing phase (1.5 hour after food) [1][2][5].

Scientists have found a significant link between This high blood sugar and several other diseases like catastrophic damage of several nerves, kidney and Renal failure, heart and vein damages, eye-sight. An uncontrolled diabetes level for a long time period can also lead a patient to dead. The growth rate of diabetes patients is enormous around the globe. From a report published in 2013, the International Diabetes Federation (IDF) predicted the probable diabetes patients around the world. It claimed that estimated about 382 million or more people worldwide are carrying excessive amount of glucose in blood and probably had been suffered from diabetes, and the report also predicted that within the year of 2035, it enormous number of diabetes patients can even exceed to 592 million. From the report of various heath surveys it is estimated about the consequent percentage of total 8.5% of the population of South-east Asian region have diabetes where about half of the population of victims even do not aware of that they are carrying diabetes silently in the body .The growth rate of diabetes patients is alarming in several middle income and emerging countries and Asian countries are the major contributor for devasting growth rate of diabetes[2].

Author: *Lecturer, Department of Computer Science, Stamford University Bangladesh. e-mail: tanvir.stamford.cse@gmail.com*

From the analysis of several health studies, it is known that the advance and predefined adequate proper knowledge about the consequences of diabetes and better and more compact prediction solution may be very effective to fight against diabetes in more convenient way and help to raise awareness among people.

World Health Organization (WHO) published a static based analysis and research-based report to focus and intensify the real diabetes scenario of Bangladesh. From the recent meta- analysis conducted by WHO reviewers showed that the recent threating pervasiveness of diabetes among Bangladeshi civilian had increased in an alarming rate, the report focused on dramatic increment sequence of growth rate from in 1995 to 2010[3]. In 1995 the rate was only 4% which increased 9% in 2006 and until the year of 2010.

According to the analysis and prediction of the International Diabetes Federation based on the analysis of several case-studies , the organization predict that the devastating rate will grow further and will be increased about 13% by the year of 2030[3].

By reviewing the previous documentations and reviewing several journals, it was confirmed that there is a serious limitation in the field of diabetes research and predictive system because there are no available suitable documentations or studies dedicated for this specific region. The prime drawback of Previous studies were that the previous models were not designed properly and combination of attributes were not properly designed. Again, the previous studies were bounded to specific region or focused on specific sex or gender.

The prime emphasis of the study is to obtain a full set of co-relating factors which are the prime responsible attributes for diabetes and to establish a predictive model to predict and identify diabetes at an early age. This model is best specially optimized for the south Asian counties like Bangladesh because to conduct this study a lot of matters and factors regarding for the specific region were taken on consideration based on need and expert opinion.

Therefore, there was a serious need for a specific model. According to expert opinion and WHO's Report guideline, the primary goal was to identify each individual, household and related fixed or specified community factors associated with the conditions. WHO found and expressed a significant connectivity or relationship with diabetes and age. In most common term, older /middle aged people have the more chance to get attached by diabetes because age is one of the prime differentiators for diabetes.

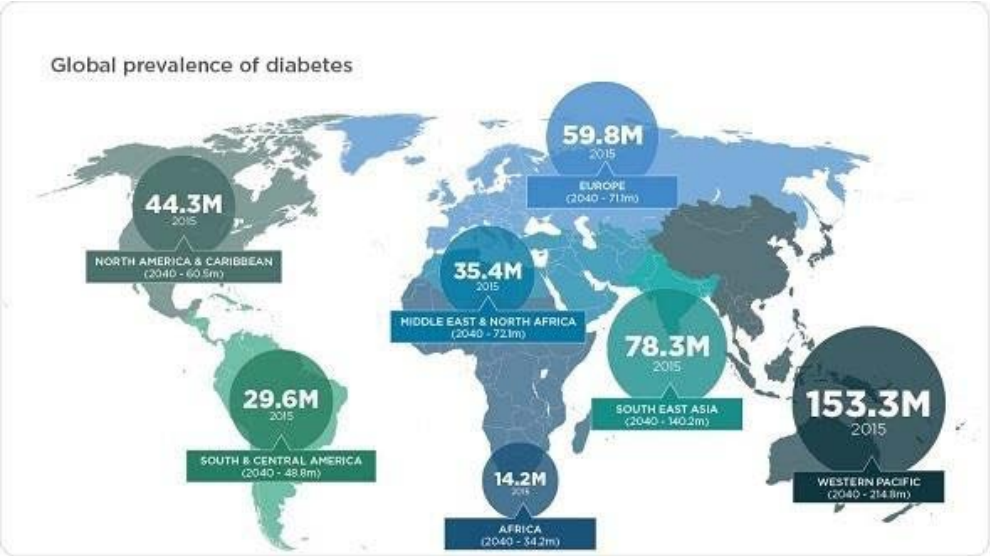


Fig. 1.1: Diabetes ratio around the world

From the result of previous study, it was confirmed that most of the diabetic affected population likely (40%) are enough educated, have sufficient knowledge about diabetes and they belong to middle class income level, where almost 13% participants were from lower income family. From the previous study, it was known that about 40% of total diabetes population were receiving regular medical check-up and proper healthcare system.

Diabetes is known as a silent carrier and it is a carrier of several deadly diseases which can cause long term health hazards. To maintain a good health score, it is needed to identify diabetes at primary stage and maintain a proper diet and exercise chart. Again, the devastating growth rate of diabetes in this region, is very harmful for the human resource management for the country as diabetes patients become unable to overtake heavy and handy task[7][9].

So, the method of prediction of diabetes at an early stage is very important and beneficial for the community. There is no available preventive methods of totally cure the diabetes and root out the disease from

the body, but there is a well-defined solution to control the glycemic index and sugar of the body.

Again, By using several data-mining techniques, it is possible to predict the disease far early and assist doctors and healthcare providers to reach in a better diseases management procedure. Analysists and researchers Patient will also get food and exercise recommendation through this system.

b) Problem Definition

Diabetes is a wide spread disease and it has some common symptoms and attributes. Family history, Age, Sex, BMI, blood pressure etc were taken into consideration to make a proper evaluation of model. The normal measurement level of diabetes is fall between to 6.0 mmol/L during the time of fasting and it will cross the level of 7.8 mmol/L after 2 hours meal.

Diabetes has 2 different types which include type 1 diabetes and type 2 diabetes. Diabetes has some specific symptoms. These symptoms appear to the people, especially those with patients with type 2 diabetes, some of these symptoms may appear lately. For the type 1 diabetes patients the symptoms may appear quickly and more severe [4].

Some common symbol of type 1 and type 2 diabetes are [4]:

- Increased amount of extreme thirst
- Increased amount of hunger [7][12].
- Sudden weight loss
- Frequently a chemical substance named ketones is found in the urine
- More Frequent and unexplained urination [5]
- Decrement of vision gradually
- Get attacked by more and Frequent common infections, like skin infections and infection in several sensitive organs [8].

From the result of various surveys and analysis, it was confirmed that Type 1 diabetes can develop in the body at any indigenous period or age, though it often found in childhood stage. where Type 2 diabetes is more widely spread to the middle-aged person and common in people older than 40 [4], though type-2 diabetes can also appear at an early age. Diabetic diseases is classified into four category. Therefore, patient can have these type of diabetes. These are given below:

Type 1 diabetes

The prime cause of type 1 diabetes is still unknown today. As per scientific documentations, combination of genetic susceptibility and environmental factors are considered as the primary reason of Type-1 diabetes .In this type, the immune system of the body surprisingly misunderstands and destroys the insulin-producing cells in the pancreas. This action is hazardous for metabolic system because in this case there is only a little or no insulin found in the body which is insignificant for metabolism. As a result, metabolism and energy transmission to the body cell is hampered and extensive level of sugar found in bloodstream [4].

Type 2 diabetes

As per scientific documentations, in type 2 diabetes, body cells become resistant to the action of insulin. one the other hand, the organ named pancreas cannot produce sufficient insulin for the body [4]. For a type-2 Diabetes, adequate exercise and proper diet plan is needed to manage the proper blood sugar level. Again, for some specific case, doctors recommend insulin to some specific patients. Doctors and health scientists around the world indicated that genetic factors, nature and environment plays an important role for creating diabetes. Overweight and diabetes Type-2 has also strong relation.

Gestational diabetes

During the period of pregnancy, the placenta produces some dedicated hormones. These specified hormones act against insulin [4]. Generally, in all the cases and types of diabetes, patients pancreas releases extra-more insulin to control and manage the diabetes. But in some cases pancreas cannot emit extra insulin which causes gestational diabetes [4].

c) Overview of the thesis

The responsible reacting factors were marked and identified based on several important factors and co-related relationship. The findings and the results of Several statics and previous studies were extensively used to make the proper evolution of the model. Then important information and findings from the results of previous successful case studies were identified and sorted for future use. Attributes were selected with extensive care and based on their contributions for developing diabetes in body. The expert opinions and doctors advice enlisted in several health journals were taken into consideration to select the proper attributes. Again, participants were classified and divided into several groups and different categories based on research demand. Several statistical evolution and informative data were the prime source of data and patients real time data depending on the Complete list of foods and different meal plan [0]. Based on the collected samples and evaluations of different attributes of each and individual patients, the desired calorie need was taken into consideration based on patients need and health need. If the diabetes patients follow the recommended diet chart and adequate exercise then it will be beneficial to control diabetes.

d) Scope of the thesis

The thesis was done based on the evolution of several related attributes and their contribution towards the thesis. The findings and recommendations of thesis will pave the way to find out a more prominent and trustworthy solution for the diabetes patients to

digenesis the disease far earlier than it appear and it will recommend a optimum lifestyle needed for the diabetes patients. The lifestyle, food habit, exercise time and taken insulin can be stored in a database for further analysis. Again, based on the comprehensive analysis and extensive data analysis, prediction of risk factor of diabetes can be calculated by using our developed model. Risk factors and the probability of patient's get attacked by diabetes can be predicted accurately to fight against diabetes in more convenient way.

e) Objective of the thesis

The primary goal of thesis is to identify the risk of diabetes at an early age to create awareness among the future diabetes patients and to manage the diabetes in much more pre-planned and organized way to fight against diabetes. As diabetes is a permanent diseases and there is no available solution to up-root diabetes from the body. The proper and well defined safety regulations can be ensured by regular assessments. Diabetic patient's recommended and optimum lifestyle was also suggested in this evaluation. It will also pave the way to reach in a compact solution by predicting diabetes earlier and upcoming diabetes victims will become more conscious about their habit and lifestyle to minimize the risk and to manage a better health index.

f) Organization of the thesis

Despite of the advancement of medical science, there is no permanent cure Diabetes and it is growing at an alarming rate which is cautious for the economical development of the country. Most of the times, the patient's health condition becomes worse of Diabetic Patients because they are often ignorant of risk factor and do not maintain a proper diet chart. So, the advance prediction of risk factor of diabetes can be life saving for the ignorant patients.

In the previous study, Many scientists used various kinds of machine learning techniques. Researchers around the world experimented and used several different types of classification algorithms. Several statistical techniques and mechanisms were used to predict diabetes in advance. Again, Doctors and Health experts also analyzed the performance of different algorithms and cross-validated the model. The previous approaches paved the way to reach in a compact solution.

In this paper, a custom designed well defined model along with a refined formula was proposed to identify diabetes at an early age and manage the proper health index in much more convenient way.

A well structured and efficient dataset was collected from the combination of various medium like Internet open data sources, survey results and questionnaires and all the data was stored in a integrated database to use it for further validation and future development of the model.

Extensive Synthetic and analysis of collected dataset was done based on the combination of various attribute. In this research, it was the primary goal to find out a proper relationship status and proportional or inverse-proportional relationship between the each and every reacting attributes. Then, the scientific evaluation and proper cross-validation process were done to recheck highest accuracy of the model.

II. Literature Review

a) Correlative Factors on Type 2 Diabetes Prevention Efforts of the Senior High School Students in Makassar

This study has found out the most common and successfully indicated related probable factors responsible for diabetes. The primary goal of the paper was to analyze associated factors and clauses related to DM specially for the level of teenage student in the city of Masakkar, Indonesia. In this study, the primary dataset was collected from high school students and age between 11-16 year students were highlighted and focused to determine the diabetes Meletus's impacts and related reasons to analyze the risks and the threads of diabetes for the teenager and to prevent it at an early stage. Data was collected in various methods like questionnaire, survey etc. The study is based on Indonesia, where DM prevalence is estimated increase from 8.4 million to 21.3 million between 2000 and 2030. The devastating growth rate diabetes affected patients is hazardous and it is indicating a highly health disaster. By analyzing the datasheet thoroughly, it was found that only 6 respondents (24.0%) had parent with lower or less standard education level where majority of respondents 189 in numbers had parent with high education level, consciousness in prevention efforts on Diabetes. In this paper Researchers noticed a common fact that among the participants of the study, those parent have higher or sufficient level of education, their children are more aware about the risk factors and they are adopting better prevention efforts to protect and fight against Diabetes. This study suggest that those people with low education level had "1.27 times" at risk of suffering DM than people with high education level From this survey, it was established that high incidence number of DM type 2 because of low of parental education level on prevention of DM type 2 incidence. Again, By analyzing the datasets, case studies and reports of several health sheets, it was seen that parental support plays vital role to maintain the good-heath index (standard) and it is a key promotor for creating prevention effort against Type-2 Diabetes among the teenager group. The result of the several data analysis and data sorting techniques showed a significant relationship between parental support with prevention efforts of DM type 2 among senior high school students. The study strongly found a link of The

peer support and DM type 2 diabetes. The prime clause of the peer support was referred as providing relevant information, preventive measures on DM type 2.

In this paper, Authors observed and noted a significant relationship between hereditary parental educational awareness and health consciousness level, the strong relation of benefits perceived, barriers perceived, knowledge, peer support and social awareness and proper informational advantage can create a huge improvements and significant progress of prevention of DM at an early age

b) *Designing Technological Interventions for Patients with Discordant Chronic Comorbidities and Type-2 Diabetes*

In this present decade it is often found that Patients with Discordant-Chronic Comorbidities (DCCs) are likely to attacked by multiple complex DCCs with a set of fully contradictory medicinal requirement, prescription and guidance. This problem is disastrous because a medical professional should minimize the medication as per priority of diseases .So, there was a great demand for a help assistant system which can prioritize based on patient's health index and suggest a optimum solution for patients simultaneously. As the part of the model, authors focused on developing and publish a mobile application to evaluate the risk assessments scores on demand. The prime purpose of the application is to suggest and provide proper medication guidance based on the need and physical condition of every individual patients. The suggested application gathers a ton of useful health-information, heath report, health index etc data from every individual user, then analyses data and enable patients to assist their conditions and treatments. It's often found in the medical data analysis that the Chronic conditions and apparent symptoms last for five or more months such as common diseases like Diabetes, Arthritis, or Depression, are becoming increasingly common in patients. Due to habitat and integrated nature of diabetes and it's typical conditions, patients are asked to play an active role in their treatments schedule and planning. From the health summary and results of several surveys, it is easily understood that the patients who do not follow or maintain the standard life guidance recommended for diabetes patients are at a greater risk .It's often found that the specific fact that, The development of Discordant Chronic Comorbidities with multiple chronic conditions, has become most common and often can be seen with highest rate co-connectivity which creates difficulties for healthcare assistants and desired patients when it comes to term of managing and controlling the impact of the managing conditions. To control and manage the state of diabetes, some plethora of available tools, apps, sensing devices and various sensor tools only support the care and proper management of diabetes diseases. In previous study, it is known the proper diet sheet and diet management is the key factor to manage the proper status of diabetes. From the factbook, it is known that The prime challenge in studying patients with comorbidities arises from their compounding health factors and health assessments issues, which states often leads the affected patients lead to more sicker and more spending time in enrollment of hospital admission. This is the primary barrier of understanding the proper guidance of self-management assessments of their diseases and it's associated risk factors. Based on interviews conducted with patients with Type 2 Diabetes and other Discordant Chronic Comorbidities, researchers designed a mobile application based on the barriers patients faced in successfully managing their treatment as well as some of the solutions they used or wished to use. The overall goal of this mobile application is to encourage patients to inbound in the application assessment exercise to improve their long-term health and quality of life.

By Approaching forward on this certain topics, Researchers emphasized and tried to develop this application and participate in testing with users with the ultimate hope of releasing this application to the general public. In addition, Researchers are extensively looking to find out the optimum ways to manage diabetes in a more convenient way with the supervision of computer intelligence.

c) *Recurrent Neural Networks with Non-Sequential Data to Predict Hospital Readmission of Diabetic Patients*

It's recognized that Hospital readmissions and vulnerable health index rates are the indicators of poor quality of Medicare, such as inadequate discharge planning and care coordination. It's often consider that the frequent readmission and lower health index can be avoided by certain methods and propositions. In this paper, a Recurrent Neural Network model is carefully designed to predict whether a patient would be readmitted in the hospital or his/her health index parameters will be reevaluate to gain the highest productivity with several machine learning algorithm .IN THIS Study, it is found that RNN showed highest prediction precision to target high risk patients and prevent recursive admissions. Hospital readmission and degradation of Health index what will happen when a patient within a specified time interval or timeframe, who had been released from a hospital with vital increment of health condition is admitted again. Again, a lot of research studies and publications proved that healthcare centers can be engage in several activities like clarifying patient discharge instructions, coordinating with patient's health conditional index, handling with post-acute care providers, vibrant cleaning mechanism to reduce the rate of readmissions of patients. In is paper, therefore raises a big question that which patient groups or which type of patients must be targeted to effectively reuse and redesign available

resources for preventing readmission and to use the classified information for special case study. Many predictive and specially designed Models that can predict accurately these are of a great help for hospitals all over the world as they can put extra efforts on high risk patients and can decrease their readmission rates.

In this experimental procedure of Research topic, the prime motto was to redesign, analysis and construct a powerful model to predict exact numbers of diagnosis's measurements and different type of machine learning Approaches and models were used to predict with highest accuracy. In this case, especially Recurrent oriented Neural Network outperformed the rest of the machine learning models in the prediction quality in the scale of productivity and accuracy. The knowledge, experimental results and outputs gained from the journal can effectively improve the traditional health system to target high risks patients, reduce rate of readmission and deliver better health care.

d) *Development of Indian Weighted Diabetic Risk Score (IWDRS) using Machine Learning Techniques for Type-2 Diabetes*

Medical experts and scientists have expressed their opinion that detection of diabetes at an early phase can be a lifesaving effort. Advance Diabetes relating factors and different screening tools such as Diabetes Risk Score (DRS) can effectively assist diagenesis and detecting diabetes accurately and help to prevent the diabetes among pre-diabetes phase at an early time before diabetes occurs. In current evaluations and assessments, Researchers have observed certain related issues in the available data and advocate the need to address the same. In this paper it's established a novel South-Asian regional Weighted Diabetic Risk assessments and co-relating factors. Different Machine Learning algorithms such as distance based clustering with Euclidean distance, k- means etc techniques were used by the researchers as a part of establishing a profound diabetes risk assessment tools to analyze the contribution of associated factors like blood pressure, age, stress and life quality BMI, diet, physical activity to boost up high plasma glucose level. In this paper ,the strategy to establish a strong and co-relating relationship between several differentiating factors , establishing a formula and then test and validates the formula with several test datasets to ensure the maximum accuracy. On an research World health organization referred that South-Asian countries citizen's are affecting on diabetes on this last two decades encounters at an devastating rate due to several depending factors. In this paper, the researcher collected datasets from various data sources, conducted surveys and used previously available data and information's to represent informational support. Data is collected form the south-Asian populations mostly from Bangladesh under the supervision of medical professionals. Several collected and trustworthy datasets were also used to strengthen the decision. Different types of Machine learning algorithms and advance data sorting principles are used for determining threshold values for various parameters when it was needed. A proper diabetes evaluation system or function is calculated for each factor like BMI, age, phenotypes, personal medical history, family history, diet, physical activity, stress and life quality. The genetic property, phenotype, lifestyle, working habit and some others factors are seriously related to diabetes. Different individual research, case studies and scientific studies have been proposed earlier by scientists to reduce the risk of diabetes to reduce the risks of diabetes, it's needed to differentiate the relationship between diabetes and different co-relating factors to fight against the risks of diabetes at an very early stage.

In this study, several reacting mechanisms, techniques and elements were successfully sorted which is very important to bring a new dimension in healthcare imagining prediction system. Different type of surveys, questionnaire, data synthesis techniques and computational intelligence were successfully used to identify and analyze the risk factors and their scores.

e) *Study of Type 2 Diabetes Risk Factors Using Neural Network For Thai People and Tuning Neural Network Parameters*

Advanced datamining techniques and analyzing tools are very Efficient to detect and predict diseases and their relating risk factors at an early age. In this paper, Researchers are trying to find out the relating factors which are mainly responsible for Type-2 Diabetes and proposed a relating solution to identify diabetes. In this paper a complete set of related factors which includes blood pressure, weight, body mass index, family history are considered as a primary factors. Again, smoking and alcohol consumption were considered as a strong co-relational factor based on their linked found in several researchers. To analyze and synthesis data BNN algorithm was as used. To collect datasets and sample information for training set about two-thousand samples of various health attributes were managed from BMC Hospital, Thailand. Based on previous learnings and previous research suggestions ,this paper found a strong relationship status and divided the risk level in there consequent stages i) low risk denoted as -1 point, ii) Medium denoted as between the range of -1 to 1 point and iii)High Risk was marked as the cautious level and contributed a single (1) scoring point for each risk based on different scale of measurement (unit) depending on the weighted contribution of linked factors like Family history, Age, Sex, BMI, blood pressure to make a proper evaluation of model. It was also added 1 point to the risk score for smokers and consumer of alcohol with timeframe of 4 weeks or more to summarize higher risk capability. By

analyzing the documentations of the paper it established a U-shaped relationship with the consumption alcoholic drinks and smoking habit. The major findings, research analysis and conclusions was divided in two different portions. In this study, authors Initially identified the major related and responsible factors and made a complete a list of the proper co-relating factors based on the evolution of collected datasets and previous records. Then , the all concerned factors and related terms were intelligently sorted and divided into three sophisticated categories based on their of different level of contribution to diabetes. Atlast, the study was concentrated on acquiring the learning rate with the tuning of BNN parameters.

This study concentrated in some vital factors and redefined the traditional reasoning methodologies which provides a better performance markup, higher accuracy level and better predictability compared to existing solutions and predictive analysis. From the result analysis of the paper, it was summarized that The prediction accuracy of the proposed strategy was not as good as expected, but in this paper, authors focused on the best optimum strategy to find out a better solution in future to predict diseases in much smarter way.

f) Data-Based Identification of Prediction Models for Glucose

From Result of various Health surveys and analysis, it is known that Diabetes mellitus is one of the wide spread diseases in all over the world .There are many co-effective factors which mostly responsible for the appearance of Diabetes, but there is a general or common reason between every single diabetes patients is that they might have deficiency in insulin production or insulin is not functioning well to improve the digestive system . It's advised to all the DM patients to track the regular status of blood glucose to maintain a proper control of the glucose count in the blood to become healthy and active. In this paper, it was observed that common barrier to control the diabetes or glucose level by a semiautomatic model is to monitor the mechanism of glucose levels in blood interact with insulin, diet intake or other factors interact with each other .In this paper, a set of traditional and classical identification techniques such as Holt's smoothing, classical simple smoothing model was compare to genetic programming models and techniques to evaluate the working efficiency of the model. Again, to maintain a proper and balanced autonomous glycemic control, a glucose control and blood sugar level monitoring principles and algorithms is extensively needed to outperform all existing solutions. In this paper, Authors put main emphasis to develop a forecasting or predicting model to the evaluate the level ricks DM based on trustable parameters like the real-time measurement of blood glucose. The Researcher also tried to predict the real-time basics blood glucose monitoring system and this algorithm would successfully measure the blood glucose level on the real time, it will analyze all the details and classified data and refer an insulin inhibitor system to supply the necessary amount of insulin particles based on the patient's need and health condition on the real time. The researchers have collected tons of data and heavily analyzed the data in terms of the space direction and the power spec-trum. for the 10 in-silico patients.

In this study, the researchers have reached in a conclusion that the combine package of both the previous Grammatical evaluation model and genetic programming is the best suitable techniques to predict, identify and manage the issue. This proposed approach will bring a new revolution and new strategy to adopt with next generation diagenesis and prediction modules to predict and fight against diseases at an early stage.

g) Improve Computer-Aided Diagnosis with Machine Learning Techniques Using Undiagnosed Samples

Now-a-days, different types of computer aided diagnostic tools, various predication and machine learning algorithms are used to identify the root causes and responsible factors. Again, to predict the risk of several fatal diseases in far advance and several computer aided tools and gadgets are extensive used today to assist human to prevent diseases more effectively or to maintain a good health score. To analyze and to understand thoroughly about a certain disease usually a huge number of diagnostic samples, opinions, surveys etc are needed to be collected, examined and analyzed to sort out the effective responsible factors and it's impossible for expert to analyze, simplify, synthesis this vast amount of information. That's why authors of the paper put emphasis to develop a new technique to analyze data faster. In this study Researchers proposed a effective semi- supervised machine learning algorithms named Co-Forest. Researched marked the new algorithm as an extended and extensively modified version of existing machine learning algorithm named "Random Forest". This algorithm is better for providing the analysis result and giving final hypothesis assessments compactly. [0].

Semi-supervised learning combines the both labeled and unlabeled data to extensively synthesis and extract the required information to establish a reliable and trustworthy hypothesis. The study suggests that, To plan or design a conventional methodology from scratch, the desired "co-training" data should be described by two sufficient and redundant attribute subsets.[0]In this methodology, each of the section of classification- division must be independent or act like as independent attribute and will capable of providing sufficient scopes unique learning capability. In this paper, author denoted L as a tag of labeled set and U denote unlabeled set.

In this co-training mechanism, 2 different sets of classifiers are trained from Labeled data, after that circumstance, each of sets should selects the most confident contents in Unleveled data to label from its point of View[0].

This study extensively focused on the usability of unlabeled data to boost the extensive learning capability, train from the unlabeled data and to save a lot of time in the field of health science. This approach is revolutionary and it will bring more pace in sample data management process, comparing and analyzing a ton of information in a short range of time frame.

h) Diabetes Prediction Using Ensemble Perceptron Algorithm

Today's people food habit is largely dependent on ready-made, high sugar and high calorie enriched foods. Medical experts and health scientist's advice the every suspected or affected diabetic affected person to diagnosis the level of glucose in blood in a routine cycle, which is costly and time consuming. The extensive use of data mining and machine learning algorithm with the assistance of computer aided system can effective be used to predict, identify and maintain diabetes in a controlled manner way. In this paper, authors proposed a whole new machine learning methodology and mechanism which will effectively predict the risk of diabetes for the unidentified patient and the working procedure of the new algorithm was tested on 3 different datasets to ensure the effectiveness. Several A broad range of machine learning algorithms, data mining tools and specially designed computer guided equipment are now effectively used to analyze medical data and to reach in a medical solution for any specific diseases. In this paper, researches pave the new effective way to successfully diagnosis of disease in a most convenient, compact and more rapid way. Several and different type of customized Machine learning algorithm is now vastly used to analyze medical data and to reach in a medical solution for any specific diseases, In this paper, authors suggested a new type called "Ensemble Perceptron Algorithm (EPA)" is proposed. This profound attention marked on the algorithm because this methodology is used to utilize the classified method of Perceptron Algorithm method of unseen data by a new proposed method with the help of Boosting algorithm.[0].

In this paper, Authors divided the working principle of the proposed method into 2 consequent phases. At the session of training stage, a broad range of collected samples recognized as the training set are analyzed by the perception algorithm in the cycle of arbitrary iterations and the help of packet algorithm and the cycle of the iteration will come to and an end after identifying the best weight vector. At last, the discovered weighted vector is kept in an array for further use. Then, by analyzing the weight vector, the profound analysis, score and remarks of training sets then data and scores will be reevaluated and extensively calculated by using a described function which was discovered in the paper [0]. Based on the extensive findings on several different domains, the prime factors responsible for DM were placed according to the descending order for further use .In this paper, authors considered "positive" for those resulted values which are greater than zero and rest of the values are referred as negative. After all, the analyzed sample elements are need to be properly labeled and separated by desired divisions as per the analysis of results achieved from the tests [0]. The prime approach of The Machine learning algorithms is that it stores informational attributes of several participants for medical survey and then analyze the data heavily to prepare to construct a model. In this study, the researchers identified the key factors based on the proof of certain medical evidences and then suggested a profound relationship with diabetes and it's associated risk factors. It is expected that, The learning and relational data gathered from the proposed model can effectively be used in near future with certain modifications for medical prediction of undiagnostic patients to accurately identification of the risk of the disease.

i) Prediction of diabetes based on personal lifestyle indications

Diabetes Mellitus develops in the body when there are higher or uncontrolled blood glucose level exists in the blood-plasma cell for a long time. Recently, Researchers noticed that an uncontrolled level diabetes for a long period of time can cause serious health hazards including blindness, kidney and renal failure to the affected patients who do not maintain a standard pro-diabetic lifestyle. In this study, it was marked that diabetes has a keen relationship with a person's attitude, lifestyle form factors. That's why the authors of the paper greatly devoted to establish a profound and strong relationship status between diabetes and it's associated risk factors like (age, Blood pressure, sex, Body mass index, waist circumference etc) and put their emphasis to develop a model. In this study, various algorithms like a Chi-Squared Test of Independence and another data analyzing technique named the "CART" (Classification and Regression Trees) were applied to test and analyze data. To integrate this proposed model with computer based Data clustering system, the proper cross validation steps of the process needed to be performed to ensure quality. From the analysis of previous study and research work, it was identified that the people in the age margin of 45 years or above, having high blood pressure, BMI range beyond the 25 and having a common genetic history of diabetes are the most vulnerable participants to be considered and if the participants do not follow the proper diet chart or do not take proper physical exercise (minimum 40 minutes /day) having these described attributes, these group of

people have the highest probability to fall in diabetes in the near future. To conduct the research work and to build a relationship model, Authors of the paper collected the primary data about various relationship parameters like (age, BMI, BP, sex, sleep time, Exercise time) from various sources by surveys, questionnaires and categorized and leveled the data in several bounds based on the research requirement. In this study, it is found that for the categorical dataset, an algorithm name "CART "prediction model performed the accuracy level of 75%.

Again, In this paper Researchers have investigated the collected datasets and found that High blood pressure and unbalanced diet habit and consumption of junk food have a deep relationship with diabetes and this assumption and profound relationship will bring a new era in healthcare diagenesis.

j) Diabetes prediction using Medical Data

Dr. D. Asir Antony Gnana Sin [1] in their research they presented a diabetes prediction system based on some existing algorithms like Naive Bayes (NB), function-based multilayer perceptron (MLP), decision tree-based random forests (RF). Some specified and custom techniques as well as some well specified algorithms were used to find out a brand new and effective concept of new machine learning techniques and learning to bring out a whole new process of diagnosis of diabetes in advance. Then this model was tested with different testing methods such as 10- fold cross validation (FCV) and furthermore use percentage split with 66% (PS), and use training dataset (UTD) to check the accuracy of the system. Some effective concepts-processing techniques were used by the authors to increase the overall prediction precision level of the proposed model. They concluded that the pre-processing technique produces better average accuracy for NB compared to other machine learning algorithm. They gave the diabetes datasets into the machine algorithm (NB, RF, MLP) and noted the accuracy with different test methods (FCV, UTD, PS). Then for removing the irrelevant feature through the pre-process the dataset is given into the correlation-based feature selection. This is a looping process. They used WEKA software and collected datasets from University of California, Irvine (UCI) machine learning repository.

This proposed approach and the learnings from the study will definitely bring a new revolution and brand new effective strategy to adopt with next generation diagenesis and prediction modules to predict and fight against diseases at an early stage.

k) Prediction on Diabetes Using Data mining Approach

Pardha Repalli et al. [2] in their research they predict how likely the different group of aged people are being affected with diabetes based on their life style and for finding out factors responsible for the individual to be diabetic. In this paper, authors considered some statistical datasets and information. Based on the learnings from the datasets, some specialized data sorting techniques were used based on demand in order to understand which group of aged people are being affected by this disease.

To establish a structure of the model and to find the co-relative factors , two algorithmic techniques were used to predict accurately .They are i) binary target variable decision trees and ii) regression models . The best model is selected by running multiple models such as step wise regression , forward regression , back ward regression, decision tree with entropy. They have used the dataset of 50784 records with 37 variables.

Variable selection method was used by the Researchers to identify the target (input) variables for the study. High Blood Pressure, Cholesterol Last check, Heart disease, Los all teeth, Years Education etc are important input variables to predict the binary target variable. In this paper, Researcher used the parameter: age both as nominal and quotative variable. By considering various different attributes like young age, middle age and old age, authors divided and placed them in 3 separate categories. People with age above 45 years mostly affected by diabetes, they concluded. Moreover they are suggested to visit for regular checkup, dental checkup and cholesterol checkup frequently in order to control the diabetes. They also suggested young and middle age people for visiting clinic in order to check whether they have diabetes or not. Age, High blood pressure, last cholesterol check, adult BMI, Lastflu shot and heart attack are the factors that also responsible for the individual to be diabetic.

l) Predictive Analysis of Diabetic Patient Data using Machine Learning and Hadoop

Diabetes Mellitus generally referred to as Diabetes is one of the form of Non Communicable Diseases. Diabetes is so critical that it forms a long time complication situation associated with other types of diseases. For this purpose a wise and definite way have to be found to reduce the overall impact related to diabetes by doing early prediction of Diabetes patients history that can be datasets related to diabetes patients.

This paper proposed a systematic way that consists of machine learning and datasets analysis procedure includes Hadoop and map reduce approach. This methods are used to analyze the huge amount of datasets and find a pattern matching for it and also implements the missing data during analysis of data and this procedure is followed for predictive analysis. For machine learning purpose supervised machine learning approach is followed-Supervised machine learning is an approach where the overall input types and what sorts of output can be generated or what sort of output can be produced in any of the cases is previously known. For this approach it uses its previous

datasets or past experiences to trained up itself and provides an expected result.

Hadoop or Apache Hadoop is one of the open source framework which forms a computer cluster in a distributed way and it is massively used for analyzing massive amount of data in a very easy and less amount of time. For analysis and processing of further data map reducing technique is followed ,it is a way of processing data in a more reliable manner i.e this framework has the capability of processing data in a parallel and distributed way. And it is done in two phases-Firstly it will take input of data (map phase) and will convert it into intermediate data in the form of key value pairs and the next phase is the reduce phase where, by integrating and analysis of all the key values from map phase it is converted to final output.

One of the vital and major factor that is used data analysis is all the attributes that are present in datasets and used for analyzing and results obtained is used for predicting the future risk. During the dataset analysis one of the major factor that causing problem is the values that are missing of any one of the attribute i.e null values that can cause serious affects on results. So to overcome this situation classification clustering is used and by using this technique missing values are replaced with their attribute mean. For this Missing Value Imputation (MVI) algorithm is used by them. This algorithm firstly identify missing values from all attributes and then for each attribute It calculates the attribute mean. Afterwards it impute missing values in dataset with attribute mean and finally it combines missing values and datasets to produce the final result.

m) Application of Data Mining Methods in Diabetes Prediction

Medical field refers and deals with accuracy. Without accuracy in this field it can cause serious negative effects on patient.

This paper refers that early diabetes prediction can be done through the use of 5 types of Data mining techniques-GMM, SVM, Logistic regression, Elm and ANN. Among the mentioned techniques ANN (Artificial neural Networks) gives the highest accuracy rate and that result is much more closer to the actual result.ANN is a method where it's consist of multiple layers or a cubical design, here the single path traverses its way from front to back and this helps in resetting weights on the frontal neural units. ANN includes Layers and network functions. The ANN consist of or configured of three layers namely- input, hidden and output. Firstly the input layer or neuron defines all the inputs that will be given and this inputs are non other than all the attributes of the datasets. According to the paper they have used 7 attributes so their neurons is also 7. Hidden layers receives inputs from input layer and provides output to output layer. The most important work of hidden neuron is, it assigns a weight for the input neurons and this assigned weights shows the relevance and importance of particular and specific input to hidden neurons. Mathematically it can be defined as a neurons network function f(X) is a combination and composition of other function gi(x) and this can be again defined as composition of some other function. The most widely composition is the non linear weighted sum where $f(x)=k(\sum i\ wi\ gi(x))$ where K is the activation function i.e it's a predefined function .The activation function provides a small out change when a small change is made in the input. In this paper they have used ANN to predict the diabetes and the result is 0.89 which is closer to actual result and this result is obtained when the hidden layer number is 2 and hidden neuron is 5. That is it is found that by using ANN method it gives highest accuracy rate of 89%.

n) A Clinical Perspective

Diabetes is one of the common type of diseases where the blood sugar level in body become immensely high it generally of two types namely type 1 and other one is type 2.

Type 1: Type 1 is a kind of diabetes in where it is a discontinuation or disorder of glucose regulation and it is characterized by autoimmune destruction of the pancreatic beta cells that produces insulin and it leads to hyperglycemia and it have higher tendency to ketoacidosis. It is more general and seen in among children but in many case it may appear at any age. Genetic marker and the presence of antibodies can assist to identify diabetes. Antibody markers of autoimmunity that is against beta cell includes autoantibodies islet-cell and autoantibodies against insulin, decarboxylase, glutamic acid or tyrosine phosphates IA-2 and IA-2β, and ZnT8.3.Containing at least one or more than one of this are present during fasting hyperglycemia it was initially detected in persons where 85% to 90% of people can eventually contain or may develop type 1 diabetes. It is found that some patients and mostly children and adolescents contains ketoacidosis as the first symptom of this disease. In less common cases and typically in older patients, it can present with the mild fasting hyperglycemia or diminished glucose level tolerance. T1 diabetes is not a linear progression disease but it progress at a variable pace in different patients. Symptoms and sign including higher level insulin deficiency and hyperglycemia include polydipsia, fatigue, weight loss, polyphagia and polyuria. This are causing defective transport of glucose from the blood vessel/stream into body tissues and it results in increased glucose levels in the blood and moreover it elevates glucose in the urine and concomitant calorie and fluid losses with the urine. For this when insulin level falls down to such a low level lipolysis cannot be able to suppressed and products containing fat metabolism naming ketone bodies is accumulated in the blood and due to hyperventilation it

leads to metabolic acidosis and compensatory respiratory alkalosis.

o) Application of Data Mining Methods in Diabetes Prediction

In any sort of medical field the most important factor is all about accuracy. Without accuracy in this field it can cause serious negative effects on patient. So accuracy is the most important factor.

According to this paper early prediction of diabetes is made through the use of 5 types of Data mining techniques-GMM, SVM, Logistic regression, Elm and ANN. Among all the five techniques ANN (Artificial neural Networks) provides the highest rate of accuracy ANN is a method where it's consisted of multiple layers or a cubical design, here the single path traverses its way from front to back and this helps in resetting weights on the frontal neural units. ANN includes Layers and network functions. The layers are-Input layer, hidden layer, output layer. The input layer or neuron defines all the inputs that will be given and this inputs are non-other than all the attributes of the datasets. According to the paper they have used 7 attributes so their neurons is also 7 Hidden layers receives inputs from input layer and provides output to output layer. The most important work of hidden neuron is, it assigns a weight for the input neurons and this assigned weight shows the relevance and importance of particular and specific input to hidden neurons. Mathematically it can be defined as a neurons network function f(X) is a combination and composition of other function (x) and this can be again defined as composition of some other function. The most widely composition is the non linear weighted sum where $f(x) = K(\sum i\ wi\ gi(x))$ where K is the activation function i.e it's a predefined function. The most help and useful characteristic of this activation function is that it provides a small out change when a small change is made in the input. In this paper they have used ANN to predict the diabetes and the result that was assuming to be the best is 0.89 and it is obtained when the hidden layer number is 2 and hidden neuron is 5.That is it is found that by using ANN method it gives highest accuracy rate of 89%.

p) Blood pressure and ageing

Increase in blood pressure with the increasing of age can of many varied factors and it is also depended on many cases like lifestyle and living environment of different person. BP seems to be rise or fall with age. It is of two types systolic and diastolic blood pressure in short SBP and DBP. With the increase of age the blood pressure is associated mostly with the changes relating with arteries, large artery stiffens and also with increase of risk related to cardiovascular the blood pressure also rises. In case of aged person with the effect of increase of systolic and decrease of diastolic pressure related to blood there causes a risk of increasing pulse pressure that consequences in blood pressure. SBP dramatically and continuously starts to increase between the age of 30>above and in case of DBP it does not show a continuous pattern but it varies with age until fifth decade it starts to rise but suddenly starts falling at the age of 60-84. According to this paper a definite level of age is chosen for identifying the BP, in case if it is classified within different range of ages it would be much more easier to identify the provable causes of increasing or decreasing of BP.

III. Diabetic Patient Data Management and Support System

a) Introduction

The primary process of the research was to determine the principle co-relating factors and their contribution and impact toward the diabetes. To conduct the research, previous learning and knowledge base of previous health reports were considered to reach in a decision. Health information and datasets are collected from various different sources like direct questionnaires, results of conducted online surveys, previously available datasets and available heath samples of diabetes patients on various health portals and recognized health journals. Samples and essential information or heath data based on several attributes were collected from different sources from the available information of more than 450 participants of various health surveys and questionnaires. Then all the necessary information and parameters were carefully sorted and selected. After extensive sorting and filtering incomplete, less trustworthy and irrelevant information were discarded. After all, relevant information of 300 participants collected from various sources from the time period of (2011-2019) years were placed and stored in a dataset for this research purpose. This dataset was the primary information source of this research. Some principal attributes were taken into consideration. The prime attributes are age, gender, Blood pressure, height, weight, BMI, sleeping time and exercise time of each and individual patient.

The output of the research work is to build a sustainable model which is essential to predict diabetes with highest precision and detect the chance of getting diabetes in near future. This system will also suggest the optimum lifestyle and exercise suggestion to the participants

b) Diabetes patient data analysis model

The performance evaluation of a health model broadly dependent on four variables. They are Participant's real time health information, Participant's food habit, Participant's exercise sheet, Participant's medical feedback. A proper health supervision for a diabetic patient is provided by this model as this model is capable of predicting the risk of diabetes in advance and it will help the upcoming diabetes victim by providing advance alert to them. In this paper, the

attributes like BMI, height, weight, sleeping time and working hours or weekly bases exercise time, blood pressure were identified as the prime reacting factors. These attributes are the dominant factor.

Calorie intake and exercise time are also important factor for the diabetes.

Proper management of diet system and medication can treat and manage diabetes in proper way.

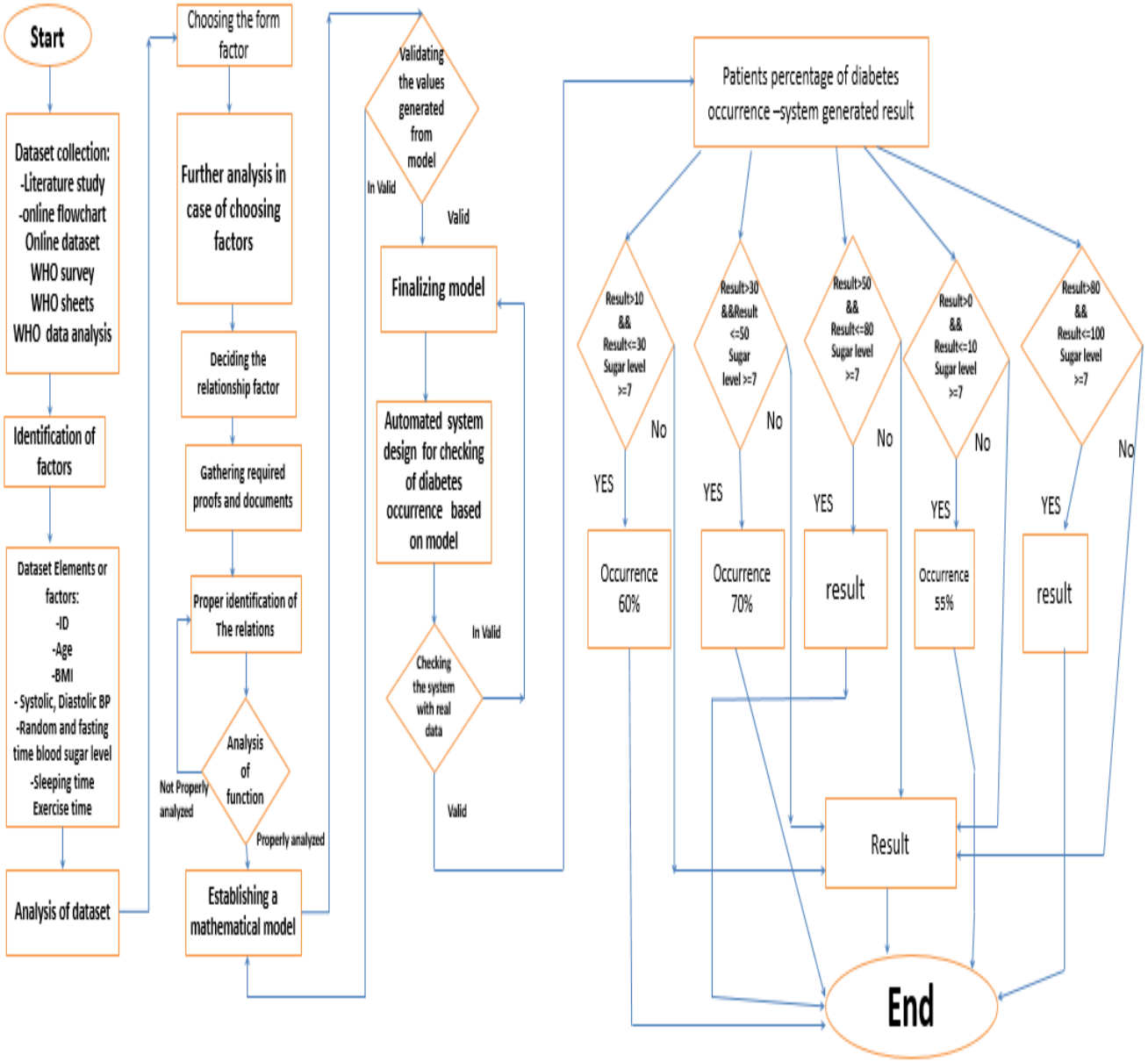
The patient containing extensive blood sugar may be suggested to take insulin.

This approach will help to manage the proper health status and control the weight, blood sugar and calorie consumption for specific patients.

c) System Architecture

Modeling of system is consisting of designing the system, processing the system architecture and integrate the proper modules and interfaces based on system requirements.

The approach and process is divided in several consequent steps. First of all, a diabetes dataset is carefully prepared and then proceed the dataset as input to the specified system analyze the data with exact precision. Then, this system is designed to perform in ready state to analyze based on input data. A well specified model and properly guided mathematical equation with proper optimization of the backend calculative format is placed in the backend of the system to analyze data. Certain terms and conditions are also set in the system to work efficiently. After taking input data from the participants, then the system measures the input data based on the developed mathematical equation. At- last, the system provides a prediction with a precise risk estimate in percentage for each individual patient. Then, this system will provide the optimum exercise goal and lifestyle for each and every individual patient. The medical experts and data scientists can use this prediction for further improved diagnosis process. However, the system is quite accurate to analyze data and to predict data for each and every individual patients. Moreover, the effectiveness of the proposed system can further be improved by.



d) Integrated Database Design

To conduct this research and to prepare the dataset, some conditions were taken into consideration. These conditions and research terms were carefully selected based on the experimental approaches and previous leanings of related research works.

In this Integrated datasheet completely emphasis on various health parameters of the patients. Again, this dataset provides a minimalistic idea about the lifestyle of the participants based on the analysis of various different parameters. By this approach, it is possible to identify the probability of diabetes at an early age.

To prepare the model realistic data was set and higher and lower bound values were carefully selected based on realistic data set, web source and medical fact data sheet. The values are carefully analyzed and not a single input in this range is out of the

Table 1 A: Attribute details list

Attribute Name	Lower bound	Upper bound
Age	1 year	123 years
BMI	10 kg/m2	50 kg/m2
Blood Sugar (Fasting)	3 mmol/L	10 mmol/L
Random Blood Sugar	5 mmol/L	30 mmol/L
Systolic Blood pressure	70 mmHg	190 mmHg

Diastolic Blood pressure	40 mmHg	99 mmHg
Exercise Time	70 min / week	2940 min/week
Sleeping time	1260 min/week	10080 min/week

In our database design we have generated following attributes. The generation process has been discussed below:

- *Participant's Age:* Patient's age is one of prime factor for this study. From the analysis of the dataset and the previous learning suggested that age has a very close relationship with diabetes. From the analysis it was observed that the people of age range belongs to 40 years to 60 years [5] have the highest risk to be get attacked by diabetes and the people of age below 40 years and above 65 years have comparatively the lower risk percentage.
- *Participant's BMI:* Body Mass Index(BMI) is an important indicator of Health index. To calculate BMI it is needed to collect Height and weight of individual patients.

  To calculate BMI it was needed to record height and weight of each and every individual participants. In this system, it was considered the existing "Guinness world records" fact book to find out the tallest and smallest heighted people's height to set the lower and upper bound of the height for the model. Though most of the participants belonged to the height range of 5 feet 3 inch to 5 feet 11inch range.
  The generalized formula to calculate BMI:
  BMI=Weight / (Height) 2

Where Weight is calculated in Kilo-gram (Kg) and Height is calculated in Meter(m). That's why, taller patients with moderate weight have likely to face less risk of getting diabetes than the shorter participants with moderate weight. To calculate BMI, it was needed to collect weight and height of the participants. In this study, it was considered the BMI range from 10 kg/m2 to 50 kg/m2 .Where the participants having the BMI range of 18.5 kg/m2 to 25 kg/m2 are considered to be healthy and participants having BMI above 30 kg/m2 are at a risk of getting diabetes in near future.

- *Participant's Gender:* Participants gender is a related factor to estimate the risk for individual patients. Patients gender is a discriminating factor for the analysis of dataset. Female patients have different type of diabetes characteristics and many women fall in temporary diabetes which is called gastrointestinal diabetes. So, Data was collected from both Male and Female participants.

Participants Blood Sugar (Fasting): Participant's blood sugar at fasting phase is likely an important indicator .It is one of the prime concern for the analyzing diabetes because patients having higher blood sugar in fasting phase likely to fall in diabetes in most of the times. By analyzing the dataset and previous study topic, it was confirmed that the participants having the fasting blood sugar range below the 3 mmol/L have the lowest.

- Possibility to fall in diabetes in near future. The Fasting blood sugar range from 3 mmol/L to 10mmol/L was taken into consideration for this system.
- *Participants Blood Sugar (Random):* Participant's blood sugar after 2 hours phase is likely an important indicator .The Random blood sugar range from 5 mmol/L to 30mmol/L was taken into consideration for this system. The random sugar should be noted with highest professionalism because any malfunctioned result or data input will change the whole result of prediction probability .The random blood sugar range above 10mmol/L is a serious indicator of getting diabetes.
- *Participants Blood Pressure:* From the analysis of several medical studies, scientists have found a significant connection of Participant's blood pressure with chance of patient's getting diabetes. The normal range of Systolic blood pressure is less than 120 mmHg and Diastolic blood pressure is less than 80 mmHg.
- *Participants Sleeping-time:* In recent studies, health scientists have found specific link to sleeping hour with the probability of getting the chance of diabetes. From the analytical reasoning of the dataset, it was found that balanced sleeping time has a inverse-proportional relationship with diabetes. The participants sleeping time were counted in hours on weekly basics.
- *Participants Exercise-time:* Exercise is the key factor to control the glucose level of the blood. Optimum exercise plan can significantly lower the blood glucose level and chances to get attacked by diabetes in near future. So, Exercise time has a inverse-proportional relationship with the blood glucose level. The participants working or exercise time were counted in hours on weekly basis.

Attribute Relationship:

Age:

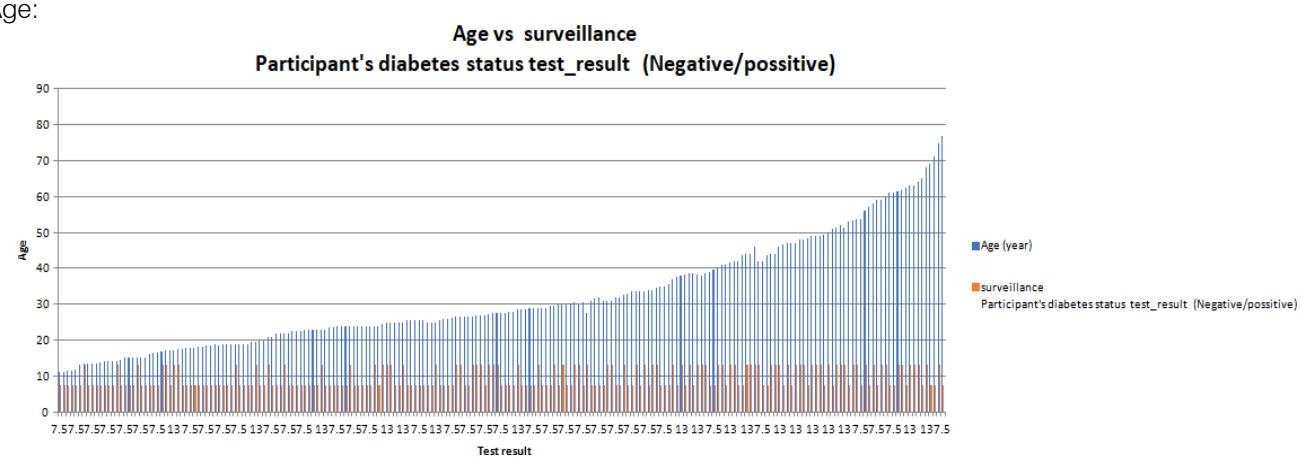


Fig. 3.2: Age vs Surveillance participants diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Age .In this dataset what was used a primary data source for the research , the Age range was between the range of 1year to 80 years of a different groups of male and female. Blue color plotted line is representing the Age (attribute). As per the information of dataset, Age started from the numerical value of 10 years old and finished at the ending point of 80 years. In this graph, Age was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 13.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that age has a proportional relationship with test_result(negative/positive).Diabetes risk occurrence is heavily linked with the age. The persons /participants under the age of 26 years have the lowest risk probability and the age range between 26-40 years have the lower possibility. The age group of above 45 years old people have the highest risk of diabetes occurrences.

From the analysis of the graph and previous studies , it is confirmed that older people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence $\propto$ participant's Age.

Diabetes Risk Occurrence = K_1*participant's Age (4.1)

Where k_1 is a constant.

Body Mass Index (BMI):

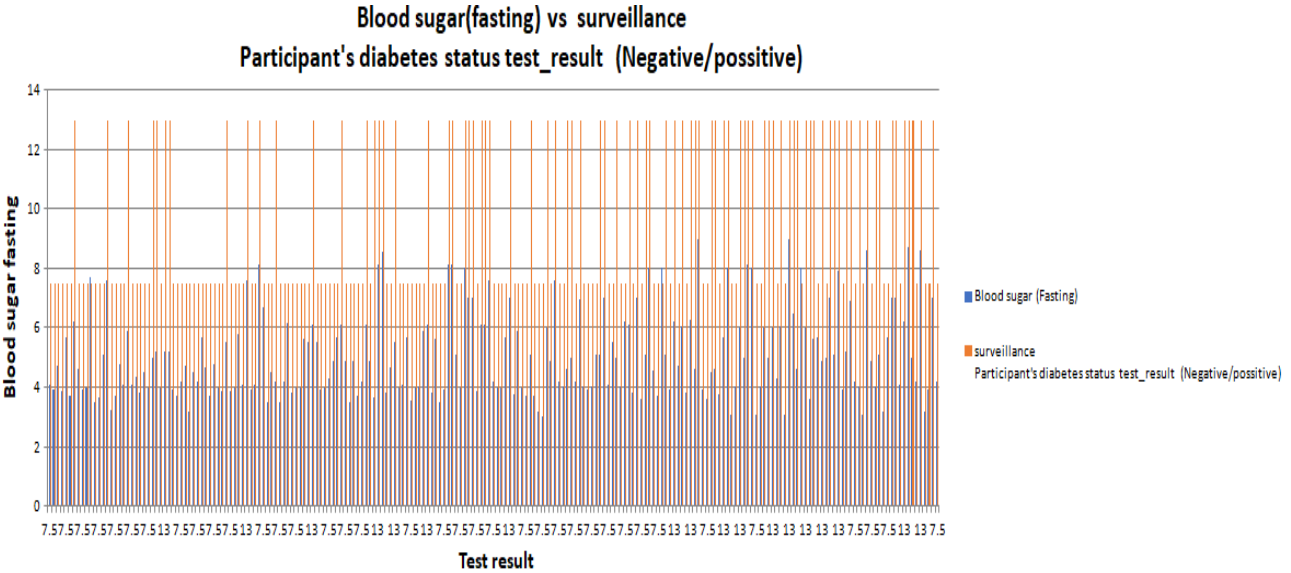


Fig. 3.3: Body Mass Index (BMI) vs Surveillance participants diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Body Bass Index(BMI) .In this dataset what was used a primary data source for the research , the Age range was between the range of 16.5 kg/m^2 to 42 16.5 kg/m^2 of a different groups of male and female. Blue color plotted line is representing the BMI (attribute). As per the information of dataset, Age started from the numerical value of kg/m^2 and finished at the ending point of 42 kg/m^2.In this graph, BMI was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "13.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value "7.0" (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that BMI has a proportional relationship with the surveillance participants diabetes status test_result (negative/ positive). Diabetes risk occurrence is heavily linked with with BMI. The persons /participants under the BMI of 21 kg/m^2 have the lowest risk probability and the BMI range between 22 kg/m^2-28 years kg/m^2 have the moderate risks. Having the BMI above 31 kg/m^2, such group people have the highest probability of risks of diabetes occurrences. From the analysis of the graph and previous studies, it is confirmed that people having the BMI >32 kg/m^2 people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ∝ participant's BMI (body mass index).

Diabetes Risk Occurrence = K_2 * participant's BMI (4.2)

Where k_2 is a constant

Blood Sugar (Fasting):

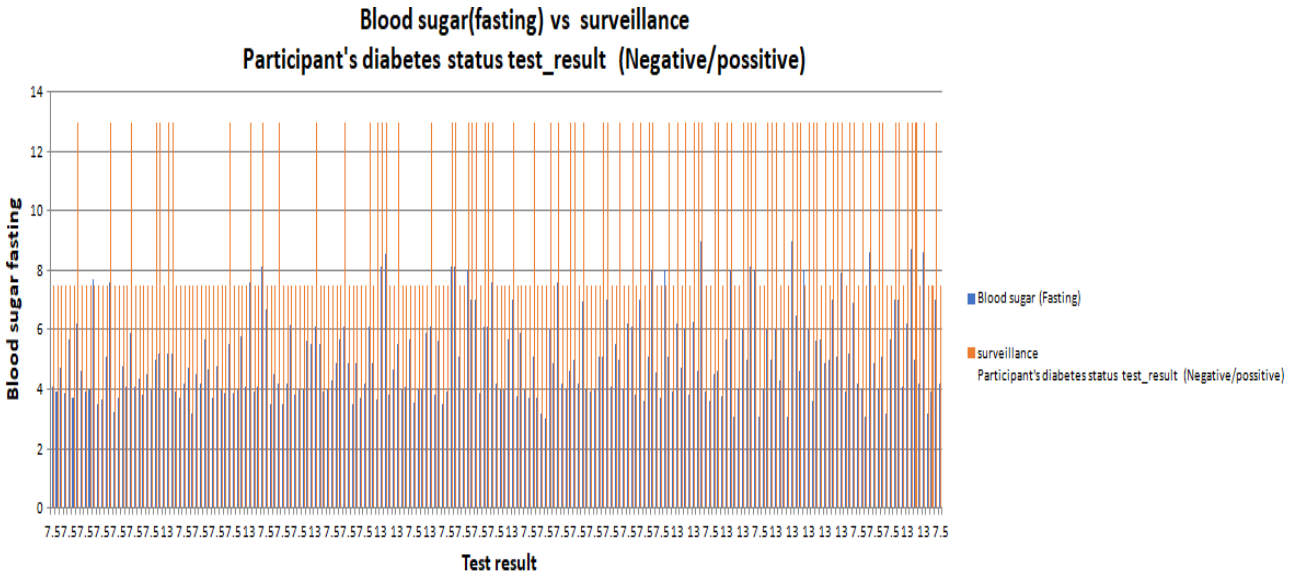


Fig. 3.4: Blood Sugar (Fasting) vs Surveillance Participants Diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Blood Sugar (Fasting) .In this dataset what was used a primary data source for the research , the Blood Sugar (Fasting) range was between the range of 3.9 mmol/L to 9.1 mmol/L of a different groups of male and female. Blue color plotted line is representing the Blood Sugar Fasting (attribute). As per the information of dataset, Blood Sugar Fasting started from the numerical value of 3.9 mmol/L and finished at the ending point of 9.1 mmol/L.In this graph, Blood Sugar Fasting was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "13.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Blood Sugar Fasting has a proportional relationship with test_result (negative/positive). Diabetes risk occurrence is seriously linked with with Blood Sugar Fasting. The persons /participants under the Blood Sugar (Fasting) range of 4.2mmol/L have the lowest risk probability and the Blood Sugar Fasting range between 4.2mmol/L to 5.1mmol/L have the moderate risks.

From the analysis of the graph and previous studies, it is confirmed that people having the Blood Sugar Fasting >6.0 mmol/L people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ∝ participant's Blood Sugar (Fasting).

Diabetes Risk Occurrence = K_3 * participant's Blood Sugar (Fasting) (4.3)

Where k_3 is a constant

Blood Sugar (Random):

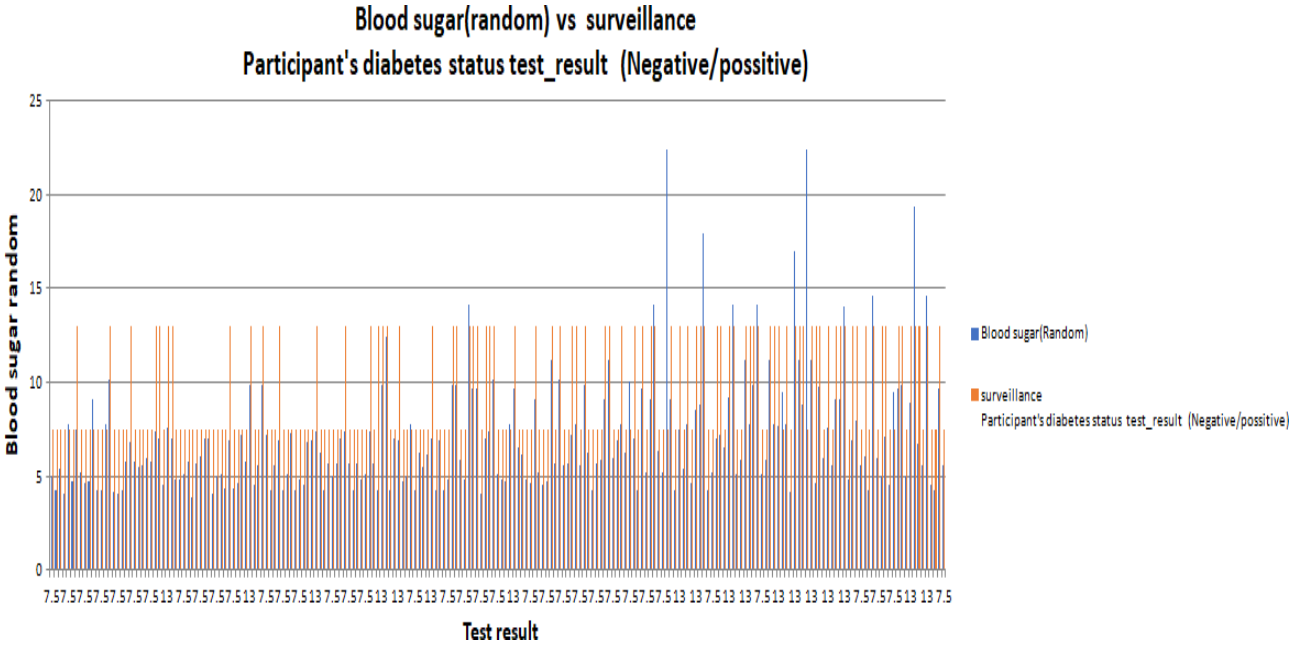


Fig. 3.5: Blood Sugar (Random) vs Surveillance Participants Diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Blood Sugar (Random). In this dataset what was used a primary data source for the research, the Blood Sugar (Random) range was between the range of 4.1 mmol/L to 18.5mmol/L of a different groups of male and female. Blue color plotted line is representing the Blood Sugar Random (attribute). As per the information of dataset, Blood Sugar Random started from the numerical value of 4.1 mmol/L and finished at the ending point of 18.5 mmol/L.In this graph, Blood Sugar Random was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 13.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value " 7.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Blood Sugar (Random) has a proportional relationship with the surveillance participants diabetes test_result(negative/positive).Diabetes risk occurrence is seriously linked with with Blood Sugar (Random). The persons/participants under the Blood Sugar (Random) range of 4.4mmol/L have the lowest risk probability and the Blood Sugar Fasting range between 5.4 mmol/L to 6.1mmol/L have the moderate risks. From the analysis of the graph and previous studies, it is confirmed that people having the Blood Sugar (Random) > 7.0 mmol/L people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ∝ participant's Blood Sugar (Random).

Diabetes Risk Occurrence = K_4 * participant's Blood Sugar (Random).--- (4.4)

Where k_4 is a constant

Diastolic Blood Pressure:

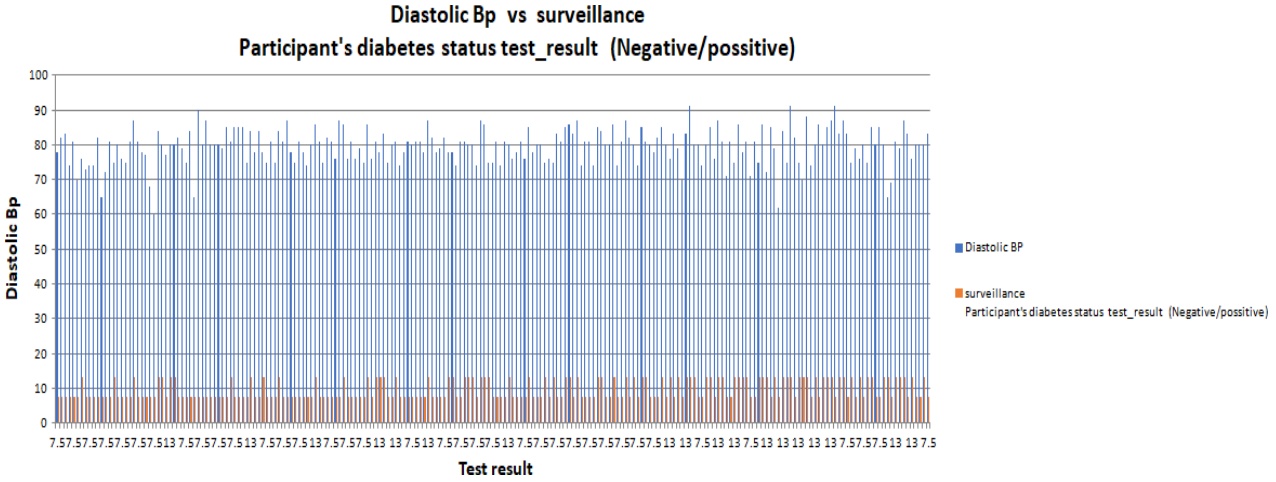


Fig. 3.6: Diastolic BP vs Surveillance Participants Diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Diastolic BP .In this dataset what was used a primary data source for the research, the Diastolic BP range was between the range of 68 mmHg to 91 mmHg of a different groups of male and female. Blue color plotted line is representing the Diastolic BP (attribute). As per the information of dataset, Diastolic BP started from the numerical value of 65 mmHg and finished at the ending point of 91 mmHg. In this graph, Diastolic BP was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "61.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value "40 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Diastolic BP has a proportional relationship with the surveillance participants test_result (negative/ positive). Diabetes risk occurrence is seriously linked with with Diastolic BP. The persons/participants under the Diastolic BP range of 70 mmHg have the lowest risk probability .From the analysis of the graph and previous studies , it is confirmed that people having the Diastolic BP > 85mmHg people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence $\propto$ participant's Diastolic Blood pressure.

Diabetes Risk Occurrence = K_5 * participant's Diastolic Blood pressure (4.5)

Where k_5 is a constant

Systolic Blood Pressure:

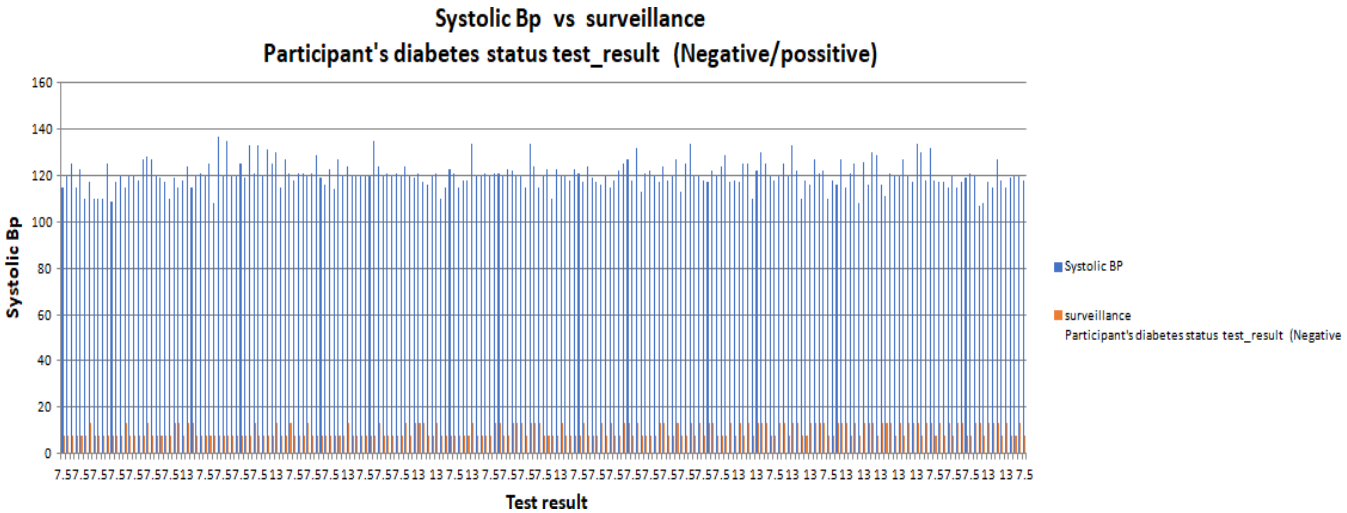


Fig. 3.7: Systolic BP vs Surveillance Participants Diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Systolic BP. In this dataset what was used a primary data source for the research, the Systolic BP range was between the range of 100 mmHg to 130 mmHg of a different groups of male and female. Blue color plotted line is representing the Systolic BP (attribute). As per the information of dataset, Systolic BP started from the numerical value of 100 mmHg and finished at the ending point of 130 mmHg. In this graph, Systolic BP was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value "61.0" (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph, for ii) Tested Negative which is denoted as numerical value "40 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Systolic BP has a proportional relationship with the surveillance participants diabetes status. Diabetes risk occurrence is seriously linked with with Systolic BP. The persons /participants under the Systolic BP range of 100 mmHg have the lowest risk probability. From the analysis of the graph and previous studies, it is confirmed that people having the Systolic BP > 125mmHg people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ∝ participant's Systolic Blood pressure.

Diabetes Risk Occurrence = K_6 * participant's Systolic blood pressure (4.6)

Where k_6 is a constant

Sleeping Time:

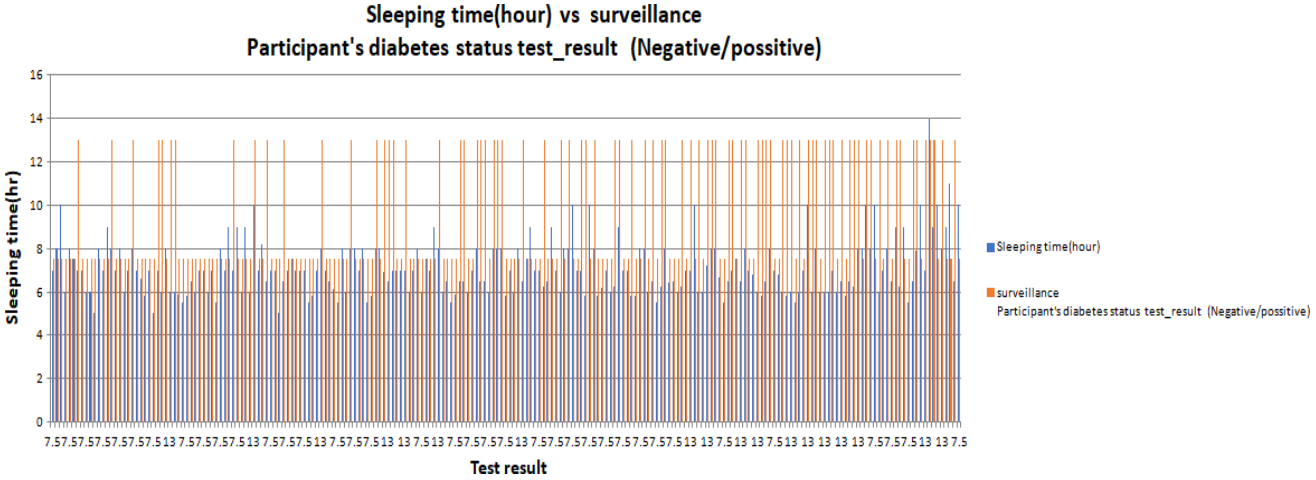


Fig. 3.8: Sleeping time vs Surveillance Participants Diabetes status test_result

In this graph ,it represents the risk of diabetes occurrence was compared with the relational attribute Sleeping time. In this dataset what was used a primary data source for the research, the Sleeping time range was between the range of 4 hours/day to 14hours/day of a different groups of male and female. Blue color plotted line is representing the Sleeping time/day (attribute). As per the information of dataset, Sleeping time started from the numerical value of 4 hours/day and finished at the ending point of 14 hours/day .In this graph, Sleeping time was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 5.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value "3.0 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph , it is clear that Sleeping time has a inversly proportional relationship with the surveillance participants diabetes status test_result (negative/positive). Diabetes risk occurrence is seriously linked with Sleeping time. The persons /participants having sleeping time range between 6 hours/day to 8 hours/day have the lowest risk probability.

From the analysis of the graph and previous studies, it is confirmed that people having the Sleeping time > 8 hours/day for a week & Sleeping time < 5.5 hours/day for a week people have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence ∝ (1 / participant's Sleeping time).

Diabetes Risk Occurrence = K_7 * (1 / participant's Sleeping time) (4.7)

Where k_7 is a constant

Exercise Time:

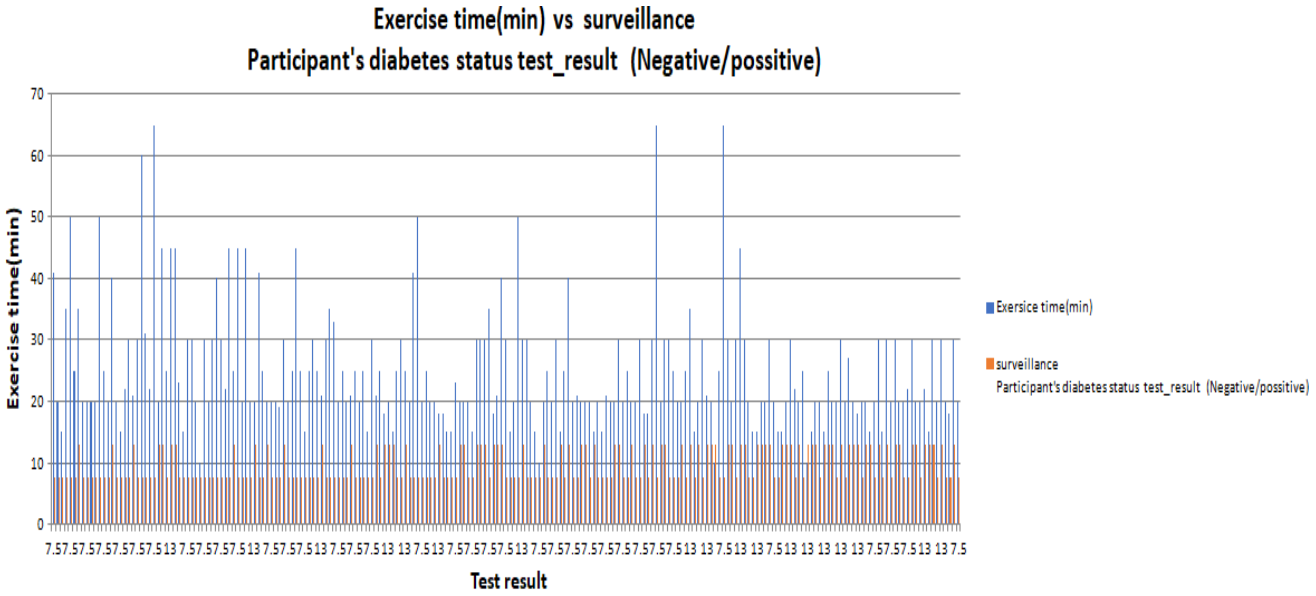


Fig. 3.9: Exercise vs Surveillance Participants Diabetes status test_result

In this graph, it represents the risk of diabetes occurrence was compared with the relational attribute Exercise time .In this dataset what was used a primary data source for the research, the Exercise time range was between the range of 10 min/day to 100 min /day of a different groups of male and female. Blue color plotted line is representing the Exercise time/day (attribute). As per the information of dataset, Exercise time started from the numerical value of 10 min/day and finished at the ending point of 100 min/day .In this graph, Exercise time was compared with surveillance participants diabetes status test_result. The test result has two different values i) Tested positive which is denoted as numerical value " 8.0 " (Yes/diabetes tested positive) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph , for ii) Tested Negative which is denoted as numerical value "3.50 " (False/diabetes tested Negative) to make and plotting the intercepting graph flatters and to make it more flexible to compare differentiating points of the graph. From the visual inspection of the graph, it is clear that Exercise time has a inversly proportional relationship with the surveillance participants diabetes status test_result (negative/positive). Diabetes risk occurrence is seriously /day to 60 min/day have the lowest risk probability .

From the analysis of the graph and previous studies, it is confirmed that people having the Exercise time < 20 min/day for a week have the higher risk of diabetes occurrences.

Diabetes Risk Occurrence $\propto$ (1 / participant's Exercise time).

Diabetes Risk Occurrence = K_8 * (1 / participant's Exercise time) (4.8)

Where k_8 is a constant

Patient's Exercise: Depending on the working type patients are given exercise time which is basically from 70 minutes to 315 minutes in a week.

Combining all the constants we get,
4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8

$$D = \frac{K*a*b*c*d*e*f}{g*h} \quad \ldots\ldots\ldots\ldots\ldots \quad 4.9$$

Where k is a constant.

By considering all the medical datasheets and references it was confirmed that diabetes risk for certain factor has a certain upper bound and risk doesn't increase exceeding that bound.

Considering The worst case scenario

Threshold age = 45 years Threshold BMI = 37 kg /m^2

Threshold diastolic Bp = 130 mmHg Threshold Systolic Bp = 90 mmHg

Threshold Fasting period blood sugar level = 6mmol/L
Threshold Random period blood sugar level = 8mmol/L
Threshold Exercise time = 15min/day

Threshold Sleeping time = 4.5hr/day minimum

So for worst case scenario all the parameters stay on threshold point. For that Occurrence will not increase after that certain point.

$$x = \frac{a*b*c*d*e*f}{2a}$$

Calculating the value of Constant K (Considering worst case scenario)

Occurrence =

$$K \frac{Age*BMI*Systolic\ Bp*Diastolic\ Bp*Blood\ sugar\ level(fasting\ period)*Blood\ sugarlevel(random\ period)}{Exercise\ Time*sleeping\ Time}$$

Here,

$$K = \frac{Occurrence*Exercise\ time*sleeping\ time}{Age*BMI*Systolic\ Bp*Diastolic\ Bp*Blood\ Sugar\ level(Fasting\ period)*Blood\ Sugar\ level(random\ period)}$$

$$= \frac{100*16200*900}{1.41912*109*37*17331.6*11998.98*(6*10^{-3})*(8*10^{-3})}$$

$$=2.78176 * 10^{-6}$$

Test case 1:
Assuming ,

Age=45 yrs = $1.41*10^9$ sec BMI= 37 kg/m^2

Diastolic Bp = 130mmHg = 17331.6 pa Systolic Bp = 90mmHg = 11998.98 pa

Blood sugar level fasting period = 6 mmol/L = $6*10^{-3}$ mol/L Blood sugar level random period = 8 mmol/L = $8*10^{-3}$ mol/L Exercise Time = 900sec

Sleeping Time = 4.5 hour = 16200 sec So calculating the occurrence

$$\text{Occurrence} = \frac{K*a*b*c*d*e*f}{g*h}$$

$$= \frac{(2.781*10^{-6})*(1.41*109)*37*17331.6*11998.98*(6*10^{-3})*(8*10^{-3})}{900*16200}$$

$$= 100$$

SO, diabetes occurrence percentage rate is 100 percent.

Algorithm: Pseudocode

Input: Participant's Age, BMI, Systolic Blood pressure, diastolic blood pressure Fasting blood sugar, Random blood sugar, working time, sleeping time.

Step 1: Collecting data from users input

Step 2: Storing inputs and passing it into assigned variables

Step 3: conversion of inputed data /parameters into standard forms and converting all into SI unit.

Age Calculation = Age*(365*24*60*60) seconds;

BMI = input value kg/m2

Fasting sugar level calculation = Fasting sugar level*0.001 mol/L;

Random sugar level Calculation = Random sugar level*0.001 mol/L;

Systolic Bp calculation= Systolic Bp *133.32 pa;

Diastolic Bp calculation = diastolic bp*133.32 pa;

Sleep time calculation = sleep time*60*60 second;

Exercise time calculation = exersice time*60 second;

Constant value K is equal to 2.78 times exponential 6;

Step 4: passing the converted values into desired variables;

Step 5: starting of calculation by using the passed variables values into the derived equation

Step 6: Analysis Report or Result is received.

Step 7: Comparing the predicted calculation with predefined sets of terms and conditions

Original result is equal to multiplication of (Constant value , Age, BMI, fasting sugar level, random sugar level, systolic and diastolic bp) which is divided by multiplication of sleep and exercise time.

1. If original result greater 10 and original result less than or equal to 30

If Random sugar level greater 0.007
Then Output 60 percent
Else output original result

2. Else if original result greater 30 and original result less than or equal to 50
If Random sugar level greater 0.007
Then Output 70 percent
Else output original result
3. Else if original result greater 50 and original result less than or equal to 80
then, Output original result
4. Else if original result greater 0 and original result less than or equal to 10
If Random sugar level greater 0.007
Then, Output 55 percent
Else output original result
5. Else if original result less than 0
If Random sugar level greater 0.007
then Output 51 percent
Else output 0.0001 percent
6. Else if original result greater 80 and original result less than or equal to 100
then, output original result
Else if original result greater 100
output 100 percent

else
Output "invalid input";

Step 8: Displaying the predicted result and risk evaluation to the user

Step 9: Ready for further analysis of different inputs

IV. Result and Analysis

For checking the Diabetes occurrence percentage rate we have used a computer programmed system which is developed according to our mathematical model. All the required attributes that we are using are taken in consideration for giving input into the system and from that we get our diabetes occurrence percentage rate. For the overall procedure 28 sets of data are given input into the system starting from age 11-77 yrs. Afterwards by using the acquired occurrence percentage rate for every individual sets of data, graphs are prepared. The graphs show the comparative analysis of diabetes occurrence rate with individual attributes. Here for every individual graph Blood sugar level(random time) and Blood sugar level(fasting time) are taken in consideration because this two attributes contributes the most crucial part for occurrence rate change because with a small change in these attributes overall occurrence rate changes at a higher or lower rate.

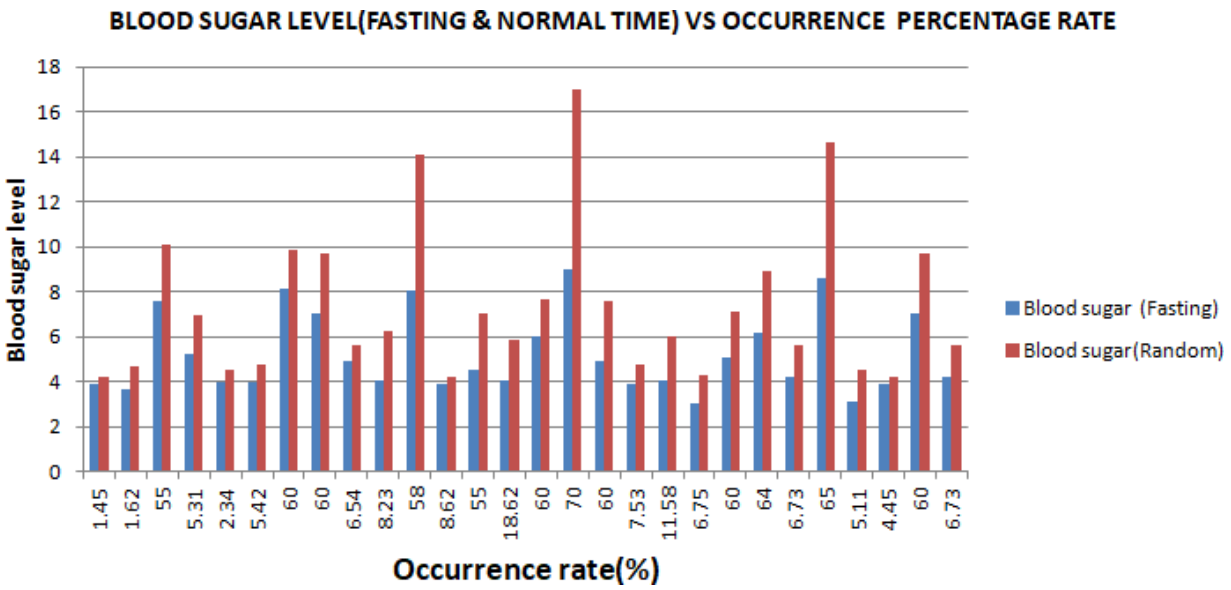


Fig. 4.1: Blood sugar level (random & fasting time) VS Occurrence percentage rate

From the above Blood sugar level (random & fasting time) VS Occurrence percentage rate graph it can be seen that with the increase in blood sugar level the diabetes occurrence percentage rate increase at a high rate. Random time sugar level has a higher effect to change in occurrence then fasting time. More importantly when the Blood sugar level exceeding the value 7 then the diabetes occurrence rate increases immensely. From the graph it can be seen with the blood sugar level 17 for random time and 9 for fasting time it gives the highest chance of diabetes occurrence (70%) and level below or close to 4 gives the lowest level of occurrence rate (1.45%).

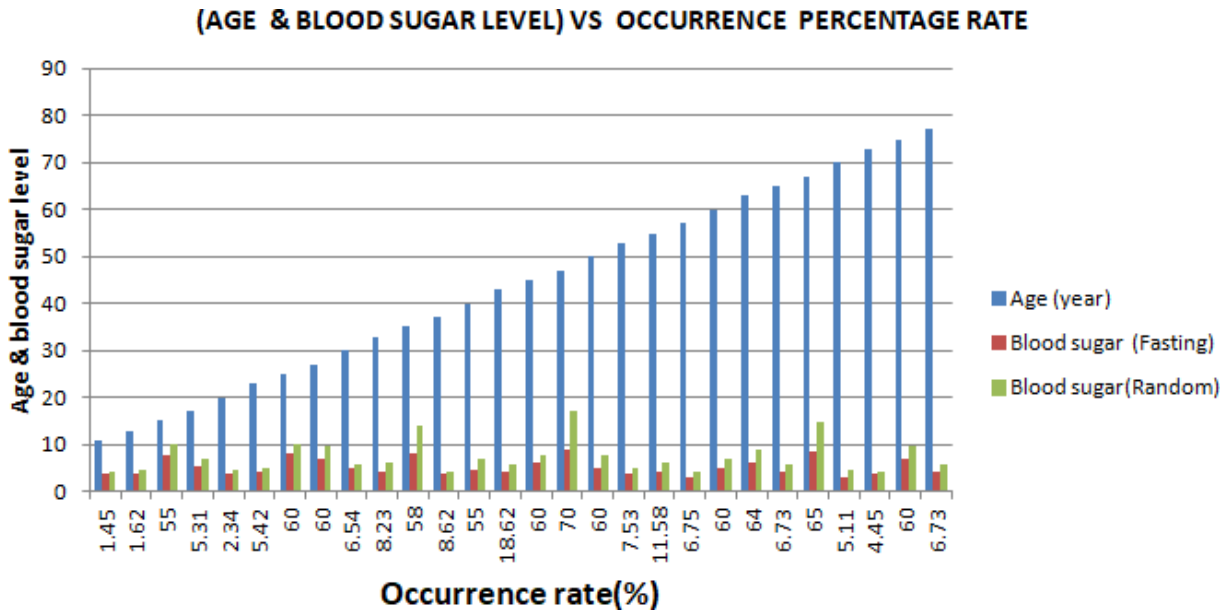


Fig. 4.2: Age (year) VS occurrence percentage rate

The above bar graph is showing the comparative analysis of Age VS Occurrence percentage rate. The graph shows that age have a very little effects on occurrence rate. Graph shows a person of age 15 have a high occurrence rate of diabetes which is 55% rather than a person of age 77 with an occurrence rate of only 6.73%.This happens because it can be seen that the person of 15 yrs of age have a higher blood sugar level then the person of age 77.

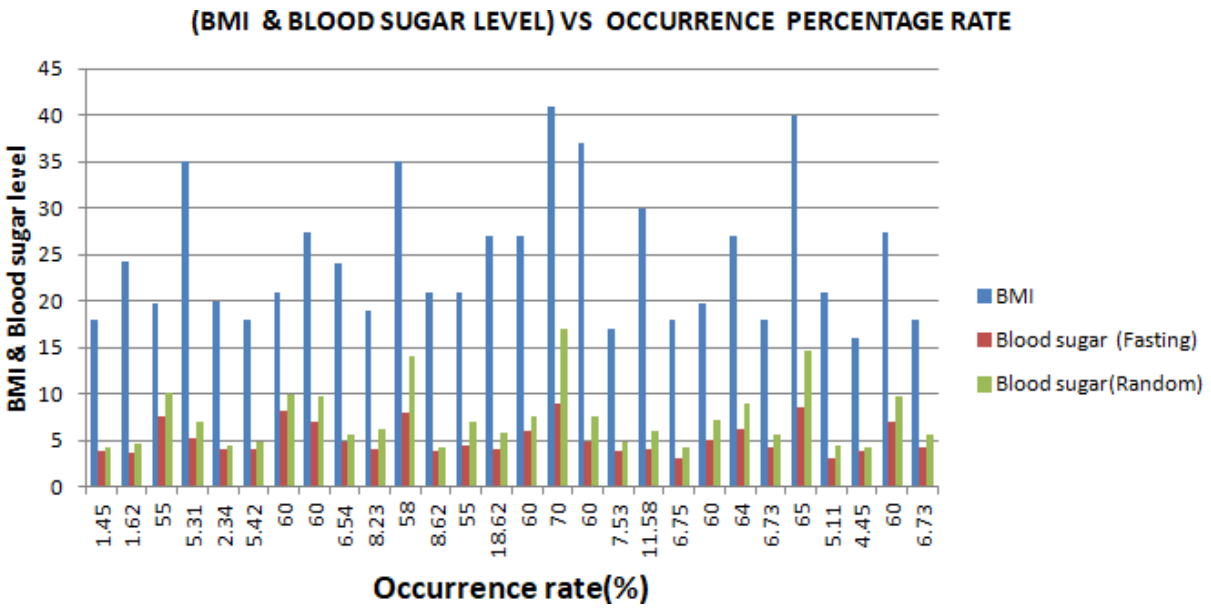


Fig. 4.3: BMI VS Occurrence percentage rate

From the above BMI VS Occurrence percentage rate graph it can be seen that BMI values those are above or very much close to 20 in presence of higher Blood sugar level have a higher rate of Diabetes occurrence rate. From the graph BMI values of 19.7,21,41,37,40 have a blood sugar level (random/ fasting period) above or equal to 7. And those values have the highest chance of diabetes Occurrence.

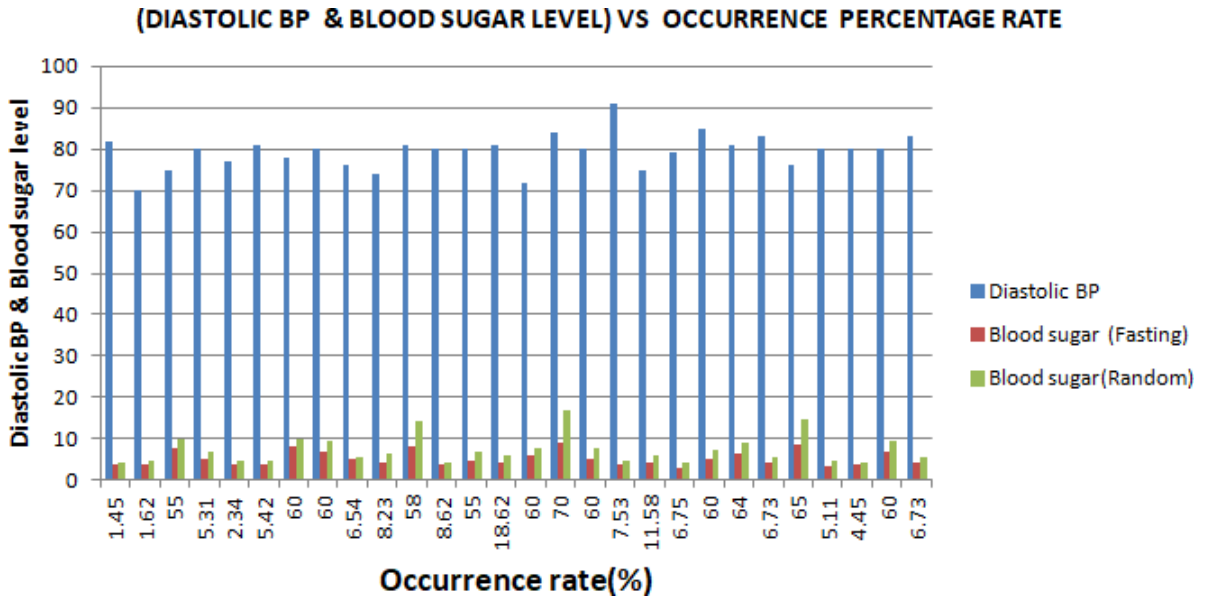


Fig. 4.4: Diastolic BP VS Occurrence percentage rate

From the above Diastolic BP VS Occurrence percentage rate graph it can be seen that Diastolic values have a little effect on the overall occurrence rate. From the graph Diastolic value of 91 have a occurrence rate of only 7.53 % but Diastolic value of 72 or 84 have higher occurrence rate(60%,70%) it is occurring due to the significant change in higher rate of blood sugar level(random) and followed by blood sugar level(fasting time).

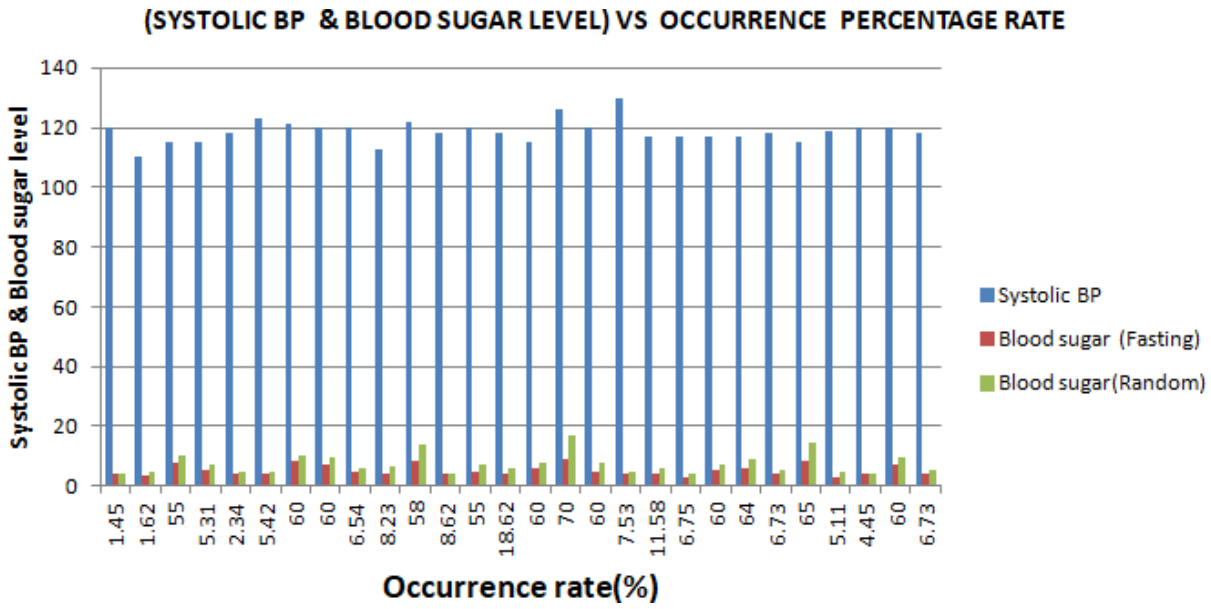


Fig. 4.5: Systolic BP VS Occurrence percentage rate

From the above Systolic BP VS Occurrence percentage rate graph it is found that Systolic values between or close to (115 to 125) with a Blood sugar level(Random/fasting) above or equal to 7 have a higher chance of diabetes occurrence.

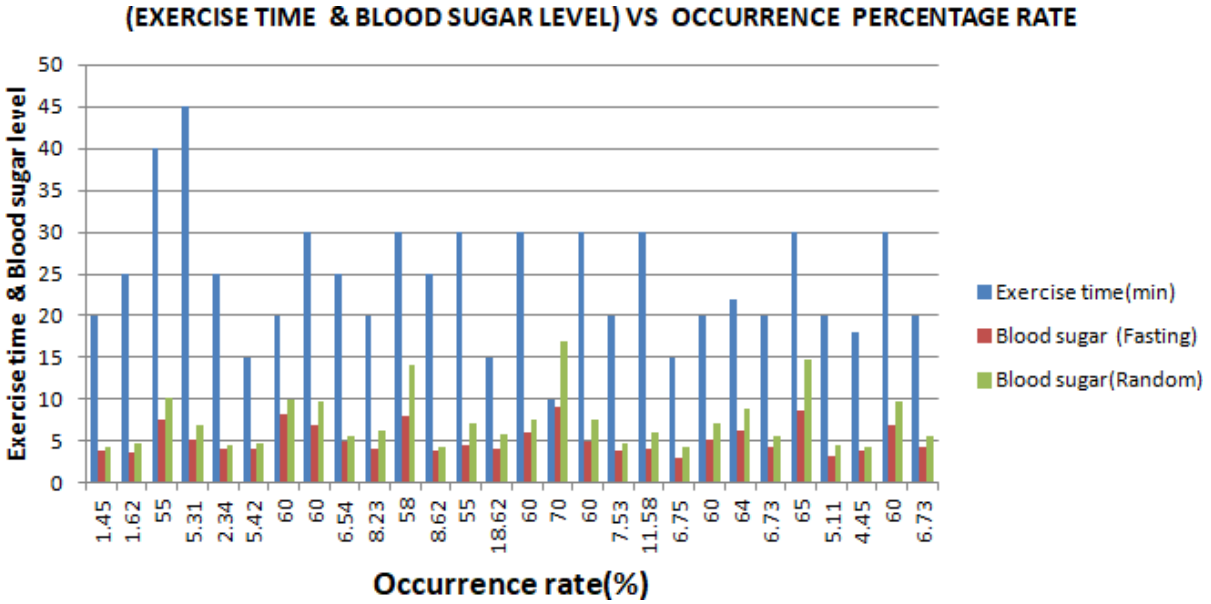


Fig. 4.6: Exercise time (min) VS Occurrence percentage rate

From the above Exercise time (min) VS Occurrence percentage rate graph it is found that with higher exercise time and blood sugar level (random/fasting) less than 7 have a less chance of diabetes occurrence and from the graph it is also found when exercise time is 40 min and blood sugar level above 7 have a occurrence rate of 55 % and again with a decrease of exercise time to 20 min with a similar blood sugar level the chance of occurrence increases by 5%. So with less exercise time and having higher blood sugar level increases the chance of diabetes occurrence.

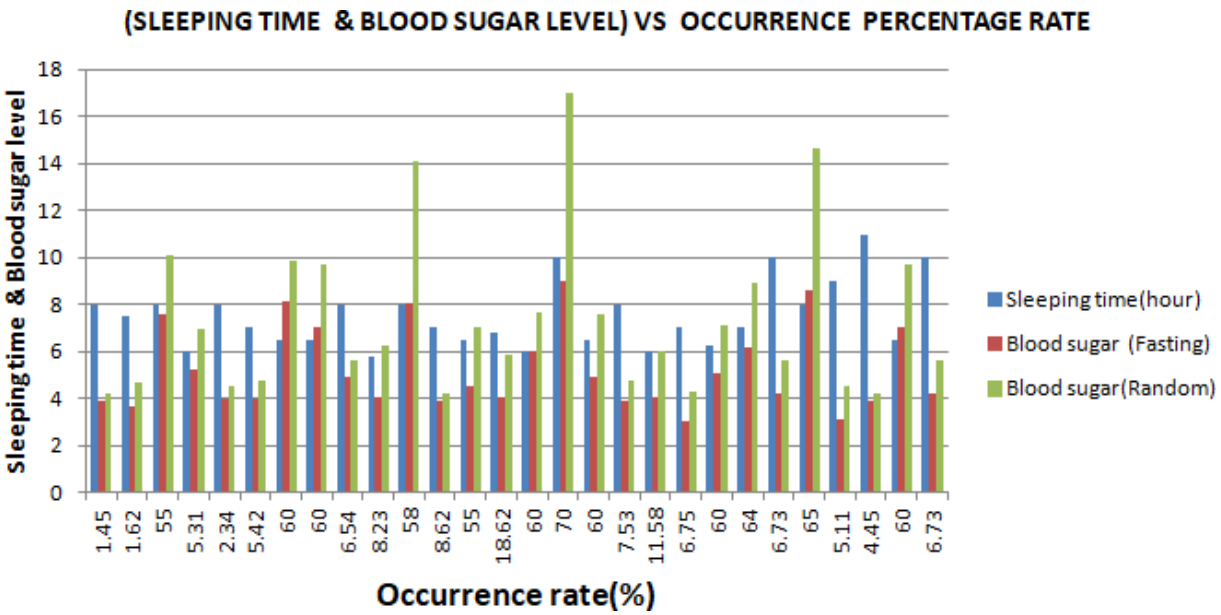


Fig. 4.7: Sleeping time (hr) VS Occurrence percentage rate

For general cases sleeping time is inversely proportional to occurrence rate that is with the increase of sleeping time diabetes occurrence chance will be decreased.

From the above graph it is seen that for sleeping time of 11hrs the diabetes occurrence rate is 4.45% but that occurs if the blood sugar level is less that if below 7. But if the blood sugar level is high sleeping time have a very little effect on the overall occurrence rate as it is found from the graph when sleeping time is 10hrs and blood sugar level(random) is 17 and blood sugar level(fasting) period is 9 in that case it has the highest chance of diabetes occurrence.

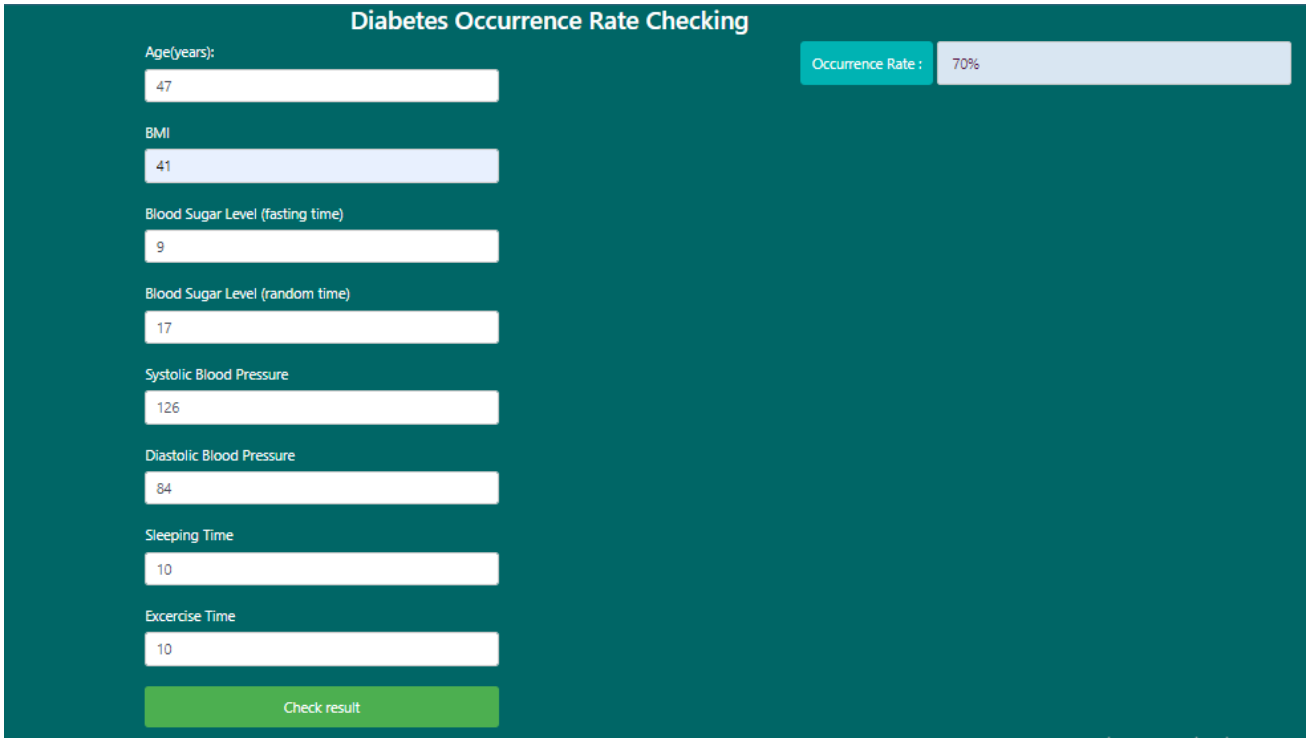


Fig. 4.8: User interface for checking diabetes occurrence percentage

In this section an overview of our developed system is shown. There are eight input fields for the user. All the inputs will be in standard unit of these parameters. Users will input their data in the text fields in proper formatting. After clicking the check result button the front end will collect the form data and will place

them in some variables and it will pass all variables to the back-end system to analyze data and generating a result. In the back- end the equation equates values and calculate the occurrence percentages of diabetes based on the mathematical model. Figure 11 shows the User Interface developed in this research to identify the occurrence of Diabetes in human body.

V. Conclusion

a) Summary of the thesis

The primary goal of the developed model is to identify the occurrence rate of diabetes at an early stage with highest precision. Therefore, to identify the crucial factors for the thesis work a largesets of attributes were taken into consideration and after extensive analysis and scientific evaluations between the attributes, some attributes were finally selected to establish a scientific based mathematical equation which is combining all the terms, co-relations and all factors in a single mathematical equation for better and fast predictability. Using machine learning and data analysis techniques, it was established that the prediction score from the developed model matches closely with previous results. The model will provide valuable result and it will be helpful to identify diabetes occurrence rate with a less amount of diagnosis time and lowest cost consumption. Though the system has some error tolerance issue but after successful experimental and testing phase, the quality of data analyzing model and software system got better and became more reliable for accurate prediction.

b) Findings of the thesis

To conduct the research work , a huge number of case studies were analyzed and 50 more related journals, health science articles, analytics, survey reports were thoroughly studied to find out the actual reason of diabetes occurrences in human body and in this paper 8 co-relational actors were indicated and their co-relationship , bindings and contribution towards the diabetes was identified and marked .Then a complete mathematical term is established based on the previous knowledge, analytical attributes synthesis and based on mathematical terminology. Established mathematical equation and concepts were combined in a single equation with a universal constant formatting All the mathematical terms were reverified in several techniques like plotting different attributes in graph to identify the correct relationship. A dataset of 250 participants of different age, groups and communities were selected for the case study and testing of the developed system. From this study it was confirmed that age is the most dominant factor and then random blood sugar level is a clear indicator of the diabetes status or diabetes level .All the attributes studied in this research like age, BMI, blood sugar, blood pressure, working and sleeping time have some contributions on diabetes risk score. A person can easily minimize the risk score by adjusting his/her life status, daily habits, food-calories intake and scaling an ideal exercise or sleeping time. Though diabetes is not preventable but the blood sugar level of any patient can easily be maintained in ideal level by inducting an ideal food, diet chart, balanced sleep and working hours and a good quality of life. The risk of diabetes will be optimized by an ideal lifestyle recommended by health nutrition experts and medical professionals.

c) Future Scope of the thesis

In this developed model, an estimated compulsion proportion between all the attributes were selected and all the attributes consist of same weighted values. Some attributes like Age and working time are primary deal breaking factors but in this work , genetics property of diabetes was not considered due to lack of proper evidence ,lack of previous studies .In future work, genetic inheritance factor will be considered for further detailed analysis .In this study, a software system is developed with manual input checking and it shows the output of risk percentage .In future work, a complete data book for every patient will be added. Interface of the computer system will be further modified. Social media's add-ons can also be added so that the system can easily fetch user data from social account for further analysis with less user input, which will become more user friendly .Our system can be also integrated with other health monitoring devices like smart watches like Apple Watch 3 or others which will be very effective to sync user data in real time basics and to store a portfolio for the patients .This system will be ready to sync data from other input sources, health devices and generate results based on the users input .Then the results will also be sent to added IoT gadgets for better health management. In the next edition our software will predict with more precision and accuracy with the extended use of IOT connected devices which will help patients to maintain an optimal lifestyle and balanced diet. In future edition, our developed software and ecosystem will also provide a better health analytic and better health management system.

References Références Referencias

1. Bum Ju Lee and Jong Yeol Kim (2014) Identification of Type 2 Diabetes RisK Factors using Pheno Type Consisting of Anthropometry and Triglycerides based on Machine Learning, 1 edn., IEEE xplorer: IEEE xplorer.
2. P. Suresh Kumar, S. Pranavi (2017) Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics, IEEE xplorer.
3. "Diabetes Fact sheet N°312". WHO. October 2013. Retrieved 25 March 2014.
4. AYUSH ANAND, DIVYA SHAKTI (September 2015) PREDICTION OF DIABETES BASED ON

PERSONAL LIFESTYLE INDICATORS, 1 edn., 2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015) Dehradun, India.
5. U. Varshney, Pervasive Healthcare Computing: EMR/EHR, Wireless and Health Monitoring, 2009.
6. JRojalina Priyadarshini, Nilamadhab Dash, Rachita Mishra (2014) A Novel approach to Predict Diabetes Mellitus using Modified Extreme Learning Machine, 1 edn., IEEE Xplorer: IEEE Xplorer.
7. P. T. Katzmarzyk, C. L. Craig, L. Gauvin, "Adiposity, physical fitness and incident diabetes: the physical activity longitudinal study," Diabetologia, vol. 50, no. 3, pp. 538–544, Mar. 2007.
8. Z. Xu, X. Qi, A. K. Dahl, W. Xu, "Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes," Diabet. Med., vol. 30, no. 6, pp. e201–e207, Jun. 2013
9. R. N. Feng, C. Zhao, C. Wang, Y. C. Niu, K. Li, F. C. Guo, S. T. Li, C. H. Sun, Y. Li, "BMI is strongly associated with hypertension, and waist circumference is strongly associated with type 2 diabetes and dyslipidemia, in northern Chinese adults," J. Epidemiol., vol. 22, no. 4, pp. 317–323, May 2012.
10. A. Berber, R. Gómez-Santos, G. Fanghänel, L. Sánchez-Reyes, "Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population," Int. J. Obes. Relat. Metab. Disord., vol. 25, no. 12, pp. 1794–1799, Dec. 2001.
11. M. B. Snijder, P. Z. Zimmet, M. Visser, J. M. Dekker, J. C. Seidell, J. E. Shaw, "Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study," Int. J. Obes. Relat. Metab. Disord., vol. 28, no. 3, pp. 402–409, Mar. 2004.
12. Adidela DR, Lavanya DG, Jaya SG, Allam AR. Application of fuzzy ID3 to predict diabetes. Intr Jrnl Adv Comp Math Sci.2012; 3(4): 541–5.
13. Affreen Ara, Dr Aftab Ara (2017) Case Study :Integrating IoT, Streaming Analytics and Machine Learning to improve Intelligent Diabetes Management System, 1 edn., International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017): IEEE.
14. R. Kumar and R. Srivastava, "Some observations on the performance of segmentation algorithms for microscopic biopsy images," in Proceedings of the International Conference on modeling and Simulation of Diffusive Processes and Applications (ICMSDPA '14),pp. 16–22, Department of Mathematics, Banaras Hindu University, Varanasi, India, October 2014.
15. T. D. Control and complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin independent diabetes mellitus. N Engl J Med, 1993.
16. AfrandP, Yazdani NM, Moetamedzadeh H, NaderiF, Panahi MS. Design and implementation of an expert clinical system for diabetes diagnosis. Global Jrnl of Sci, Engg and Tech; 2012. p. 23–31. ISSN: 2322-2441.
17. Rajesh K, Sangeetha V. Application of data mining methods and techniques for diabetes diagnosis. International Journal of Engineering and Innovative Technology (IJEIT). 2012; 2(3): 224–9.
18. S. A. Lear, M. M. Chen, J. J. Frohlich, C. L. Birmingham, "The relationship between waist circumference and metabolic risk factors: Cohorts of European and Chinese descent," Metabolism, vo. 51, no. 11, pp. 1427–1432, Nov. 2002.
19. Misra, N. K. Vikram, R. Gupta, R. M. Pandey, J. S. Wasir, V. P. Gupta, "Waist circumference cutoff points and action levels for Asian Indians for identification of abdominal obesity," Int. J. Obes. (Lond), vol. 30, no. 1, pp. 106–111, Jan. 2006.
20. G. I. Webb, Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques.", IEEE Transactions on Knowledge and Data Engineering. Aug. 2004, 16(8), pp. 980-991.
21. G. I. Webb, Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques.", IEEE Transactions on Knowledge and Data Engineering. Aug. 2004, 16(8), pp. 980-991.
22. Cash, Jill. Family Practice Guidelines (3rd ed.). Springer, 2014, p. 396.ISBN 9780826168757.
23. Saravanakumar, Eswari T, Sampath P &Lavanya S 'Predictive Methodology for Diabetic Data Analysis in Big Data', Science direct, Vol.50, 2015, pp 203-208.
24. K. Suzuki, H. Yoshida, J. Nappi, S. G. Armato, 3rd, and A. H. Dachman, "Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography," Med. Phys., vol. 35, no. 2, pp. 694-703, 2008.
25. Patel, A., et al. "Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes." The New England journal of medicine 358.24 (2008): 2560-2572.

Global Journals Guidelines Handbook 2021

www.GlobalJournals.org

Memberships

Fellows/Associates of Computer Science Research Council

FCSRC/ACSRC Memberships

Introduction

FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

Fellow of Computer Science Research Council

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.

Benefit

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Fellows can use the honored title of membership. The “FCSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 60% of sales proceeds

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

Access to Editorial Board

Become a member of the Editorial Board

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career | Credibility | Exclusive | Reputation

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Associates can use the honored title of membership. The “ACSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 30-40% of sales proceeds

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.

Associate	Fellow	Research Group	Basic
$4800 **lifetime designation**	$6800 **lifetime designation**	$12500.00 **organizational**	APC **per article**
Certificate, LoR and Momento **2** discounted publishing/year **Gradation** of Research **10** research contacts/day **1 GB** Cloud Storage **GJ** Community Access	**Certificate,** LoR and Momento **Unlimited** discounted publishing/year **Gradation** of Research **Unlimited** research contacts/day **5 GB** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**Certificates,** LoRs and Momentos **Unlimited** free publishing/year **Gradation** of Research **Unlimited** research contacts/day **Unlimited** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**GJ** Community Access

Preferred Author Guidelines

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from https://globaljournals.org/Template.zip

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

Before and during Submission

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct,* along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s') names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

Policy on Plagiarism

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures

- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

Authorship Policies

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

Preparing your Manuscript

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.

Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11'", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

a) A title which should be relevant to the theme of the paper.
b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
d) An introduction, giving fundamental background objectives.
e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
f) Results which should be presented concisely by well-designed tables and figures.
g) Suitable statistical data should also be given.
h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
j) There should be brief acknowledgments.
k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.

Format Structure

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, “What words would a source have to include to be truly valuable in a research paper?” Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.

Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

Preparation of Eletronic Figures for Publication

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

Tips for writing a good quality Computer Science Research Paper

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.

6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10.Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.

20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.

Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.

The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."

Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

The Administration Rules

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350