IOT-CLOUD-BASED HADOOP FRAMEWORK WITH CLUSTERING TECHNIQUE FOR DIABETES MELLITUS

Dr. R. PRAGALADAN
Assistant Professor & Head
Department of Computer Science
Sri Vasavi College
Erode

M. REENA IVANGLIN
Ph.D Research Scholar
Department of Computer Science
Sri Vasavi College
Erode
ivanglika@gmail.com

Abstract---**Nowadays, an ever-increasing proliferation of the Internet of Things and smart medical sensors that are fundamentally altering how healthcare is provided globally. When analyzing and applying artificial intelligence to the massive amounts of information generated by wearable sensornetworks, smart healthcare systems are frequently mde able to handle close to real applications by combining Cloud and IoT architectures. A fascinating form of computing that can accommodate massive amounts of dynamic information is cloud computing. This study concentrated on the Hadoop framework together with the clustering algorithm to address the data processing issues. Additionally, this approach makes predictions about the development of diabetes under different conditions, which is more beneficial for people. The effectiveness of two distinct environmental-friendly clustering approaches is also contrasted in this research. In this article, writers present the innovative concept of "persuasive sensing" and present promising findings from two home prototypes. The information from the home monitoring systems can also be used to create analytical models with 94% precision; it can forecast blood glucose levels for the following day.**

*Keywords---**Diabetes Mellitus; Hadoop Framework; Clustering; IoT; Cloud Computing.***

I. INTRODUCTION

Processing information and identifying insightful correlations from huge amounts of information are the goals of data mining [1]. Data is mined from a variety of computations and systems, including classification, classification, neural systems, regression, association rules, artificial intelligence, decision trees, the nearest neighbor approach, genetic algorithms, etc. [2]. Most industries now use data mining methods, including advertising, financial information processing, e-commerce, the telecommunications sector, intrusion prevention, commerce, and medical data analysis [3–4]. Medical Data Analysis is a recently developing discipline of study where specific techniques can be used to retrieve information from vast informative collections to enable effective health information assessments [5]. Since diabetes is an important health problem that impacts people all over the world, this article concentrates on data analysis related to it. According to some studies, there are 246 million diabetics worldwide, and that estimate is projected to rise to 380 billion by 2025 [6].

Using a provided data object, hierarchical cluster analysis creates a hierarchy of clusters. The recursive approaches are divided into groups according to how classified decay develops [7]. From that minute, the groups gradually converge till a cluster structure is obtained. For big data sets, the agglomerative grouping is too slow [8]. The items in a single group are split up into smaller groups, and those smallergroups are then split up even further into smaller sets. This process continues till the group configuration is obtained [9]. Linkage measures also referred to as single-link, complete- link, and median line measurements, correspondingly, characterize the lowest, greatest, and mean distances among the groups [10–11].

Recent days saw the emergence of several smarter solutions that use both the network architectures and software platforms sectors due to the maturation of IoT technology [12]. These ensure that products handle healthcare at several stages, including pediatric and senior care, managing chronic illnesses, keeping an eye on epidemic diseases, using medical cyber-physical systems, and managing personal fitness and well-being. Field sensors can be used in constant surveillance to collect context information for multisensory action recognition activities [13–14]. In an intriguing investigation, visual, mobility, and acoustic information are collected using environmental sensors, an optical-track camera, and smarter watch integrated detectors in conjunction with specialized devices for collecting physical indicators [15].

II. RELATED WORKS

In reality, this is a collaborative fog-to-cloud architecture where data pre-processing, indoor positioning, and activity recognition techniques are carried out by a home gateway while information storage for distant consumption is done in a cloud platform.

A new paradigm in research has been spurred by the anticipated expansion and present market indicators of IoT-related technologies [16]. Mobile phones are the best platform for giving feedback to diabetic patients since they are widely available, affordable, dependable, real-time, flexible, and, unlike many other innovations, are used more frequently by racial and ethnic minorities [17]. This is in addition to the development of the IoT. Mobile phones can be used as self-management tools to aid people in remembering and documenting different health-related tasks. They also make it possible for carers to immediately react to changes in a patient's health condition and evaluate continuing health tendencies.

The physiological indicators from the woman's linked collection of sensors are used in the prediction model. And the discussion of sensors delivering data eventually turns to the Internet of Things (IoT). Big data generated by the Internet of Things must be sent to cloud-based data centers [18]. Fog computing is a necessary choice to minimize latencies, which is a highly important concern in situations like healthcare [19]. FC is an addition to cloud computing that takes utilization capabilities from systems close to the edge, not a substitute for it. As a result, the FC lowers delay while also improving QoS metrics like bandwidth effectiveness and energy use [20]. This research is unique in that it offers a thorough framework for tracking pregnant women.

Although the phrases interpretability and explain ability are frequently used interchangeably by investigators and are very comparable, other publications distinguish between them. Although interpretability and explain ability have never been quantitatively defined and tested, numerous efforts have been made to describe these concepts as well as other strongly linked ideas like conciseness [21]. But none of these notions are rigorous or precise in a quantitative form. A structure called Explainable AI is applied to crack up the machine learning "black box" and make the results of the systems more understandable. Explain ability is often described as the extent to which a person can comprehend a machine learning conclusion, offer commentary on the Ml algorithm, and clarifythe reasoning behind the result.

III. METHODOLOGY

The six steps of the procedure and the associated methodologies and actions for this DSR project are shown in Figure 1 along with the general approach used in this task. Precise DSR results are also displayed in the final section. Our issue was first raised in medical literature and surveys that emphasized the significance of managing chronic diseases in general and diabetes in specific. The development of an IoT/sensor system must adhere to specific criteria. There were various actual object prototypes created as the concept was developed. The communications were created with the assistance of a medical team and were chosen to be given to patients to change behaviors. Here, the project's Institutional Review Board's authorization was secured. Following the ethical board's approval of this project, the IoT/sensor systems were installed in the two homes with diabetes elderly people patients to showcase viability. The findings were then evaluated. Every time the home installation was upgraded, we learned things that enabled us to make the design cycles better.

Fig .1. Research Approach

A. DESIGNING AN IOT/SENSOR SYSTEM

Our system's main goal was to gather diabetes-related data and identify elder individuals' behaviors in their homes. We used asensor network in our artifact. This gave the team the perfect technological framework for monitoring and reacting to physiological signals, social activity, mobility, sleep, and otheraspects that are essential to health. To change the patient's behavior, the program also sent them regular motivating and pertinent text messages. The researchers selected a Bluetooth- enabled glucometer because it had dependability difficulties and instead selected a BG monitor gadget that connects readilyto a laptop through USB and can submit BG results regularly. A wireless weight scale that communicates weight information through Bluetooth was positioned in the family room.

B. K-MEANS THE ALGORITHM

An unsupervised learning algorithm called K-means handles awell-known clustering procedure. This computation aims to utilize squared error work to reduce a goal task. The intended function isFrom the group of potential applicants that demonstrated interest. The first was a white retired man in his 82nd year thatlived in the Vista neighborhood close to San Diego. In this subject's home, our system. A 60-year-old white woman fromSan Diego who lives and works there was The second topic. She is fat and also has diabetes type

$$Y = \sum_{y=1}^{k} \sum_{x=1}^{n} \left|i_x^y - c_y\right|^2 \quad (1)$$

The grouping outcome depends on the establishment of the group emphases and their quantity, which is among the numerous merge/split choices that must be made, cluster analysis will be slower. Overall, k-means grouping is superior to hierarchical clustering for big data sets as it is more effective.

C. EVALUATION

2 and hypertension. She was regarded as a high-risk subject because of her excessive BG values. Here, a pre-/post-type of intervention was created,and the processes were put in the two homes consecutively, with an expert spending one week for each home configuring and engineering all of the IoT equipment and sensors.

The 2 weeks of the pre-study "benchmarking" phase were usedto gather all of the information without giving the participants any evaluation. The two-week pre-study was adequate to verify that the system's elements were trustworthy and stable, and it also provided evidence that the individual was carryingout the instructions. Our method

Delivered communications and comments during the post- study phase. The initial implementation took place at the male subject's residence from October 18 to November 25, 2011. The female participant underwent the second home application, which began on January 2 and ended on March 1, 2012. In this case, after the procedures, all detectors and other apparatus were taken out of the residences of the individuals shown in Figure 2.

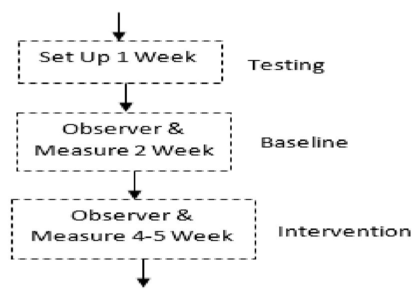


Fig.2. The Research Design for the Intervention

most important matters for k-implies clustering. K suggests that computations are valid only for statistical information. Furthermore, k-implies join to a local minimum rather than focusing on a global minimum. The comparison of k-means versus the clustering algorithm shows that as the volume of data or information grows, hierarchical clustering's efficiency degrades. Although the hierarchical algorithm has high quality, it runs slowly for various datasets. According to the current study, hypertension and the demands of the workplace both contribute to a rise in the prevalence of diabetes. This study concentrated on those characteristics for the examination of diabetics as shown in Table 1.

Table 1 Data set

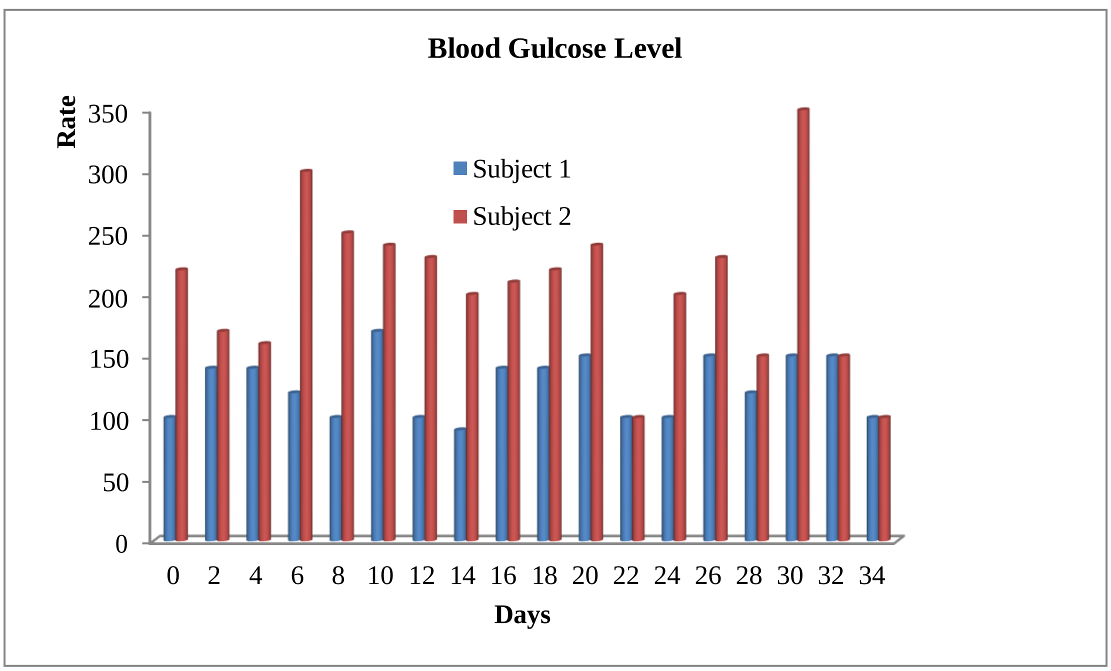
Year	Generic	BMI	Food	Hypertension	Work
2005	6.7	5.7	5	4	3
2010	6.9	7.1	7	5.5	2.8
2015	7.2	7.4	8.5	6.5	6
2020	9	9	9.6	8.5	8.9
2025	9.6	11	10.5	9.6	10.2
2030	10.7	12.2	13	13	12.3

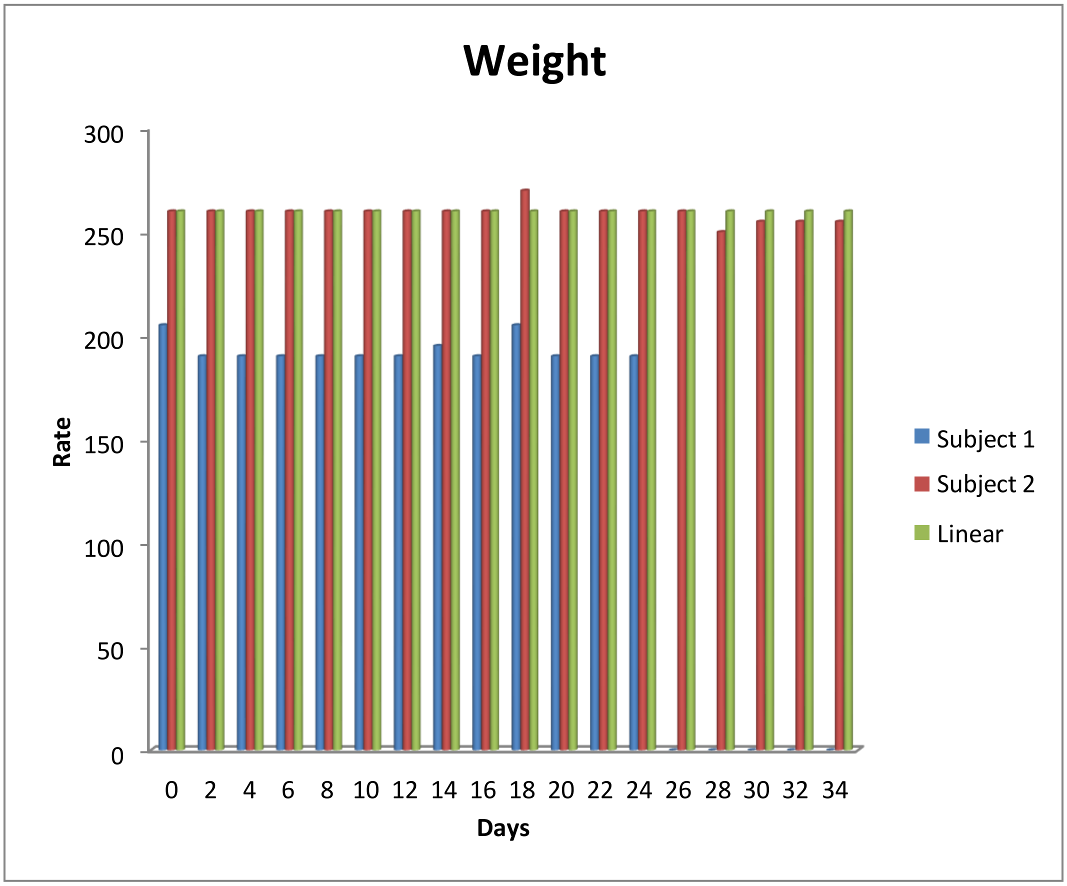
IV. OBSERVATION

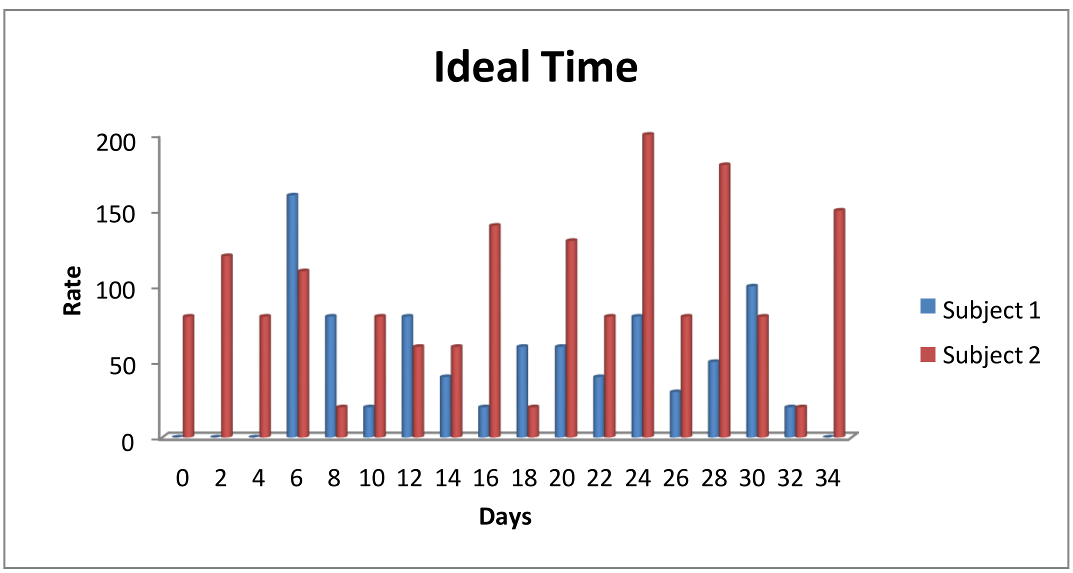
ted for BG information, see Figure 3A both participants' trend lines indicate a progressive drop. This demonstrates an increase in keeping BG levels stable. The HbA1c levels of both rising. Thus, our next Hypothesis P2 is justified.

participants were requested by the researchers both before and after the procedure. HbA1c decreased for Subject 1 from 12.8% to 6.6%, a very substantial improvement. For Participant 2, HbA1c decreased from 8.9% to 8.5%, a favorable outcome but a less marked enhancement. Subject 2's BG values fluctuate more often throughout the day. The patterns for the weight (see Figure 3B) and idle time (see Figure 3C) both point to a progressive drop. Figure 3D demonstrates an upward increase in the number of measures taken.

The purpose of our weekly email and daily text messages wasto change the participants' behavior. Furthermore, a post-intervention opinion poll was undertaken to gauge self-efficacy using a modified DSE scale. The DSE measure is a trustworthy, four-item measure that evaluates individuals' perceptions of their level of competency in managing their diabetes on their own. The survey's findings, shown in Table 2, indicate that the subjects' control over their diabetes is increasing.




Weight
Rate
Days
300
250
200
150
100
50
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
Subject 1
Subject 2
Linear


Ideal Time
Rate
Days
200
150
100
50
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
Subject 1
Subject 2

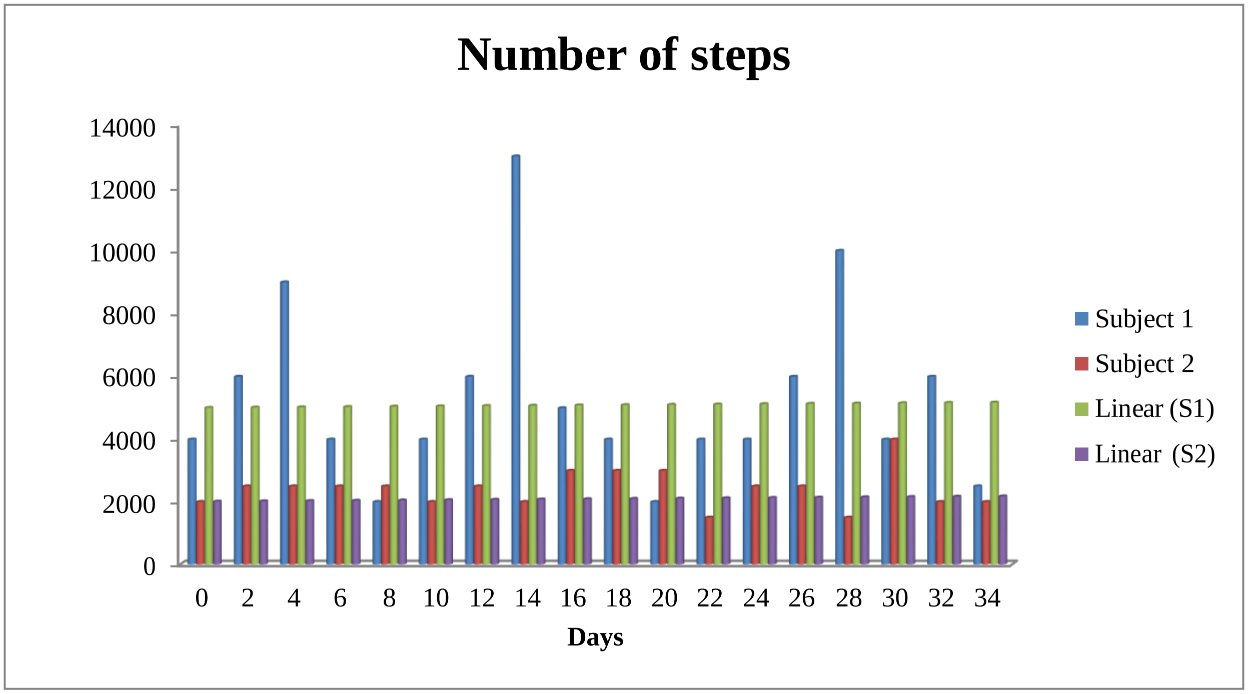


Fig.3. Intervention Results

Table.2 Diabetes Dissociative Experiences Scale

	Subject 1 (80-years old male)		Subject 2 (60-years old female)	
	Pre-persuasive sensing care	Post-persuasive sensing care	Pre-persuasive sensing care	Post-persuasive sensing care
I have faith in my ability to control my diabetes. I believe I can manage my diabetes. I am competent to handle my regular diabetic care. I am capable of overcoming the difficulty of managing my diabetes.	Agree Agree Agree Neutral	Agree Agree Strongly agree Neutral	Agree Agree Agree Neutral	Strongly agree Agree Strongly agree Agree

The pre-and post-information on several ADL variables that the systems recorded in both home installations are shown in Table 3 for comparison. See improvements in BG and weight for both individuals, as well as in sedentary time and sleep quality. The time spent in the bedroom that is not expended napping also reduces, indicating operation that is healthy for diabetic patients. It is also possible to observe that Subject 2'sdaily stage process count is lower than it was before theresearch, though this reduction gradually reverses itself throughout the interference. The consumption of calories by Subject 2 was out of her authority. How much information can be used to determine a patient's status is one of the difficulties in ADL studies. Can tracking residential ADL information help to estimate the patient's future health?

In this DSR research, we set out to achieve that goal by creating statisticalmodels as an additional DSR artifact and a result of our workTable.3.Comparison of All the Activities of Daily Living (ADL)

Rate	Subject	Weight	Blood glucose	Steps	Sleep efficiency	Lying down time	Sleeping time
Pre-study	1	203.70	157.01	2330	0.62	865	385
Post-study	2	267.80	200.02	3945	0.60	646	224
	1	194.32	126.44	5253	0.80	662	512
	2	262.80	191.43	2446	0.85	524	465

Figure 4 shows it is quite evident that the percentage of diabetics has been rising since 2010. In addition to general or persistent factors, hypertension and the nature of the job also significantly affect people's health.

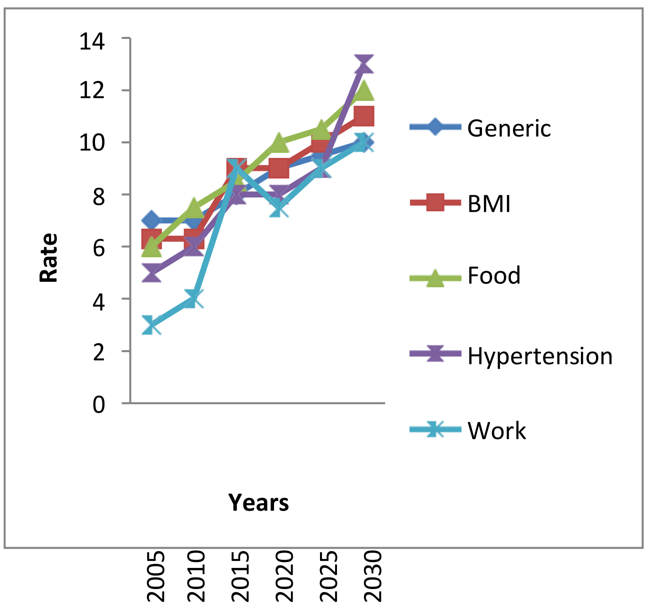


Fig. 4. Risk Factor Result

V. CONCLUSION

In this study, the efficiency, latency, and reliability of the hierarchical and k-means classification were examined. According to this investigation, the k-means clustering algorithm is more effective than the hierarchical clustering method at managing massive amounts of data in cloud computing platforms. Researchers primarily used the Hadoop framework to examine the diabetes data while taking into account factors like age, gender, and personal history. The age range of individuals under 45 to 64 was shown to have a higher rate of diabetes diagnoses. Moreover, it was also foreseen that hypertension and the type of employment had a significant impact on the overall community.

References

[1] Kumar, P. M., Lokesh, S., Varatharajan, R., Babu, G. C., &Parthasarathy, P. (2018). Cloud and IoT-based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Generation Computer Systems, 86, 527-534.

[2] El-Rashidy, N., ElSayed, N. E., El-Ghamry, A., &Talaat, F. M. (2022). Prediction of gestational diabetes based on explainable deep learning and fog computing. Soft Computing, 26(21), 11435-11450.

[3] Rodriguez, I. (2021). On the management of type 1 diabetes mellitus with IoT devices and ML techniques. arXiv preprint arXiv:2101.02409.

[4] Ali, F., El-Sappagh, S., Islam, S. R., Ali, A., Attique, M., Imran, M., &Kwak, K. S. (2021). An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Generation Computer Systems, 114, 23-43.

[5] Islam, S. R., Lloret, J., &Zikria, Y. B. (2021). Internet of Things (IoT)-based wireless health: Enabling technologies and applications. Electronics, 10(2), 148.

[6] Rodríguez-Rodríguez, I., Rodríguez, J. V., Molina-García- Pardo, J. M., Zamora-Izquierdo, M. Á., &Martínez-Inglés, M. T. (2020). A comparison of different models of glycemia dynamics for improved Type 1 diabetes mellitus management with advanced intelligent analysis in an internet of things context. Applied Sciences, 10(12), 4381.

[7] Balamurugan, K., Latchoumi, T. P., &Ezhilarasi, T. P. (2022). Wearables to Improve Efficiency, Productivity, and Safety of Operations. In Smart Manufacturing Technologies for Industry 4.0 (pp. 75-90). CRC Press

[8] Asghari, P., Rahmani, A. M., & Haj SeyyedJavadi, H. (2019). A medical monitoring scheme and health-medical service composition model in cloud-based IoT platform. Transactions on Emerging Telecommunications Technologies, 30(6), e3637.

[9] Garikapati, P., Balamurugan, K., & Latchoumi, T. P. (2022). K-means partitioning approach to predict the error observations in small datasets. International Journal of Computer Aided Engineering and Technology, 17(4), 412- 430.

[10] Priyadarshini, R., Barik, R. K., &Dubey, H. (2018). Deepfog: fog

computing-based deep neural architecture for prediction of stress types, diabetes and hypertension attacks. Computation, 6(4), 62.

[11] Abdali-Mohammadi, F., Meqdad, M. N., &Kadry, S. (2020). Development of an IoT-based and cloud-based disease prediction and diagnosis system for healthcare using machine learning algorithms. IAES International Journal of Artificial Intelligence, 9(4), 766.

[12] Lee, K. H., Urtnasan, E., Hwang, S., Lee, H. Y., Lee, J. H., Koh, S. B., &Youk, H. (2022). Concept and proof of the lifelogbigdata platform for digital healthcare and precision medicine on the cloud. Yonsei Medical Journal, 63(Suppl), S84.

[13] Latchoumi, T. P., Swathi, R., Vidyasri, P., & Balamurugan, K. (2022, March). Develop New Algorithm To Improve Safety On WMSN In Health Disease Monitoring. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 357-362). IEEE.

[14] Alfian, G., Syafrudin, M., Ijaz, M. F., Syaekhoni, M. A., Fitriyani, N. L., & Rhee, J. (2018). A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing. Sensors, 18(7), 2183.

[15] Meraj, M., Singh, S. P., Johri, P., &Quasim, M. T. (2021). An Analysis of Malaria Prediction through ML-Algorithms in Python and IoT Adoptability. Annals of the Romanian Society for Cell Biology, 25(6), 14098-14107.

Mohiyuddin, A., Javed, A. R., Chakraborty, C., Rizwan, M., Shabbir, M., &Nebhen, J. (2022). Secure cloud storage for medical IoT data using adaptive neuro-fuzzy inference system. International Journal of Fuzzy Systems, 24(2), 1203- 1215.

[16] Herder, C., Karakas, M., & Koenig, W. (2011). Biomarkers for the prediction of type 2 diabetes and cardiovascular disease. Clinical Pharmacology & Therapeutics, 90(1), 52-66.

[17] Zhang, P., Schmidt, D., White, J., &Mulvaney, S. (2018, June). Towards precision behavioral medicine with IoT: Iterative design and optimization of a self-management tool for type 1 diabetes. In 2018 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 64-74).IEEE.

[18] Rodríguez-Rodríguez, I., Rodríguez, J. V., & Campo- Valera, M. (2023). Applications of the Internet of Medical Things to Type 1 Diabetes Mellitus. Electronics, 12(3), 756.

[19] Singh, A., Dhillon, A., Kumar, N., Hossain, M. S., Muhammad, G., & Kumar, M. (2021). eDiaPredict: an ensemble-based framework for diabetes prediction. ACM Transactions on Multimidia Computing Communications and Applications, 17(2s), 1-26.

[20] Amar, J., Serino, M., Lange, C., Chabo, C., Iacovoni, J., Mondot, S., ...& DESIR Study Group. (2011). Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. Diabetologia, 54, 3055-306