

An Intelligent Evolutionary Schema on Precision Medicine for Diabetes Using Big Data Analytics



Badugu Sobhanbabu*^{ORCID}, Kodangalkar Fakeerappa Bharati^{ORCID}

Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Anantapur 515002, Andhra Pradesh, India

Corresponding Author Email: sobhanbabugec2015@gmail.com

<https://doi.org/10.18280/ria.370220>

ABSTRACT

Received: 12 February 2023

Accepted: 31 March 2023

Keywords:

Artificial Intelligence, precision medicine, big data, classification techniques

Precision medicine is the future of the Healthcare system, which is going to change the generalized medicine prescription. To establish precision medicine in the place of generalized medicine, new frameworks have to be designed that will utilize the existing enormous volumes of data. Big data is the term that represents massive volumes of data being generated through various means, including in Healthcare. In this work, a new framework for precision medicine is proposed beginning from big data, applying Classification techniques to avail reduced data and there on to achieve precision medicine with an intelligent medicine advisor. In the first phase, we evaluated the noisy big data with innovative hierarchical decision attention networks, Multi class classification using Map Reduce mechanism. This phase is experimented with a diabetes dataset. This output can be utilized by the next phase component called Intelligent Medicine advisor to get system recommended precision medicine, which can help the medical practitioner to join the Artificial Intelligence in their treatment.

1. INTRODUCTION

The amount of data being generated is growing exponentially, also leading to many challenges. In a positive perspective, more data, more accuracy for identifying the patterns. Proper pattern identification would solve many real-time problems, let it be in Healthcare, Social media analytics, Finance etc. [1, 2]. Efficient decisions can be taken by organizations based on the trends in this massive data. We can efficiently predict the future of many business aspects with big data analytics [3].

As the data gets its massive volumes, the unstructured data gives many challenges in storing, managing, processing, analyzing and visualizing the beneficial inputs and trends. This process is not very profitable with the traditional small-scale datasets. There are several stages and key components of big data. As explained by Gupta et al. [4], the Map-Reduce-based big data analytics and NO SQL management techniques help a lot to achieve business goals. There are a decent number of tools [5, 6] like Mahout, Apache Spark, Hadoop, FlinkML etc.

Big data is very helpful in the health sector too. Especially with the organization of the large health datasets, analytical capability and its technological support for achieving high-level medicine recommender systems like precision medicine. As per MedlinePlus precision medicine is "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person." This big data analytics helps in predicting more accurate treatment and prevention mechanisms for any disease. The precision medicine stands on the other side of the personalized medicine system. There are

various differences between any two individuals due to their internal resistance and habits. There are some areas like Blood fusion, where the treatment depends on the blood group of the patient. Our technical approach also alters the medication schema by providing personalized medicine. The researches [7] present the significance of precision medicine and Big Data healthcare applications [8].

Diabetes is one of the most concerning problems along with cancer or HIV in the real world. It gets worse for the patient, as they are not even permitted for certain kinds of surgeries as well [9]. The high glucose in the blood leads to early deterioration of the Kidneys, which will impact the other organs quickly. The higher level of glucose directly causes Glaucoma, blindness and alzheimers too. Being such an important problem, the medication is not uniform for all. There are several levels of Diabetes [10-12], which are classified by standard organizations like the American Association of Diabetes [13, 14].

In this regard, we are motivated to work on precision medicine for Diabetes. For which we considered a Diabetes dataset with more than one lakh entries and applied big data analytics and performed classification works. We proposed a schema to reach precision medicine in two phases. In this work, a major contribution in the first phase is presented. The organization of the paper follows this: The second section presents the Literature survey of big data in healthcare, precision medicine, and diabetes classification works. The third section presents our proposed schema. The Fourth section presents the Experimentation details followed by result analysis. The Conclusion and Future work direction are presented in the last section.

2. LITERATURE SURVEY

In this section, the concurrent research on “precision medicine” is presented in three parts. Precision medicine, Diabetes Classification work. MacEachern and Forkert [1] briefed precision medicine and its part in achieving Artificial Intelligence. This work aimed at identifying complex patterns used in precision medicine by using exploratory data analysis. This paper give a good understanding of the precision medicine concept. Yang et al. [9] utilized Hierarchical Attention Networks for Classification of Documents. This work is widely cited for the classification using Hierarchical attention networks. We have adopted their methodology in our Diabetes classification work.

Bohicchio et al. [8] has presented multidimensional data mining techniques on Big Data health care systems. The theme of this research was the Gynaecology, discussion on Prenatal and PReinatal treatment strategies and outcomes. Rao et al. [6] discussed the multidimensional mining techniques on Health care Big data, which gave motivation for our work Dash et al. [15] highlighted the significance of Healthcare Big data management in an efficient manner. Dash has presented how the proper interpretation of big data can lead toward an efficient precision medicine.

Gou and Xu [16] has presented a Multidimensional analysis of Big data in Healthcare. In their research work they highlighted the data privacy and security aspects of Big data in healthcare. Security is one of the most challenging items in the healthcare sector, as the data is sensitive for the patients under treatment. There are a couple of Blockchain implementations for precision medicine too as mentioned by Haleem et al. [17, 18]. Their evaluation can be done as mentioned by Anil and Moiz [10]. Wang, Lidong presented the Big data Healthcare management systems, its progress and technological advancements. Wang also presented the Cloud and Stream processing in the Big data healthcare domain as well.

Beam et al. [19, 20] opined that the precision medicine with Bigdata and Machine Learning algorithms are performing on par with the physicians. This is a good boost for precision medicine with system intelligence researchers. Their experimentation results give hope that technological advancements in this area will change the pattern of medication, even the medical practitioners will be trained with these tools in future.

2.1 Literature on diabetes

The American Association for Diabetes [13, 14] has presented a detailed description about various Diabetes levels. The causes of diabetes are also well explained in this paper. The diabetes characterisation by hyperglycemia, it also presented the classification (medical classification) of diabetes into Type1, and Type2 (broadly two types, which is considered in our research work as well) using the Glucose level is discussed. Solis-Herrera et al. [11] has presented the Big data analysis on the Diabetes dataset. In this research, various types of diabetes, including Type 1, 2, GDN, and Gestational diabetes are discussed. Solis also highlighted that proper classification of Diabetes will lead to proper treatment. Our work is also motivated by this challenge of efficient classification.

Alfian et al. [3] proposed a Bluetooth Low Energy-based personalized diabetes monitoring system. There are a couple

of sensors used in this work. The trained Machine learning model can predict the Blood Glucose level with decent accuracy, with which a proper diet and necessary medicine level can be given to the patient, which is the base for precision medicine. In our work, we have adapted this in the second phase of the proposed mechanism. The advent of our contribution will have a significant impact as the US Sandeago landmark on this work will arrive in the year 2022.

3. PROPOSED SCHEMA FOR PRECISION MEDICINE

As the diabetes has certain predefined stages, medication will be altered based on those stages. The overall proposed schema is in two parts. It is depicted in the following diagram(xx).

The first phase begins with analyzing enormous Big Data and leading to multidimensional classification. The second phase contains an Intelligent Precision engine whose task is currently done by the medical practitioners to decide the amount and type of medication (may be insulin or Glycomet with different mg) that is being suggested.

In this paper, we mainly focus on the first phase to arrive at the proper classification of diabetes levels using Big data analytics techniques. In this Research work, we have utilized Hadoop for big data processing and Map-Reduce algorithms, various data mining algorithms like Attention Network, Outlier based Multi-class classifiers, Decision trees etc. We have tried to consider maximum possible attributes unlike the works in the literature. Using proper Data Mining algorithms, it has arrived at the Classification of the level of Diabetes. Some brief introductions to the technologies and terminologies used are provided below, followed by the Proposed schema.

3.1 Introduction to Hadoop, map reduce

Hadoop is an Open software framework developed at the Apache Software Foundation (ASF) for distributed storage and processing of big data sets across clusters of commodity hardware. Hadoop was originally designed for web-scale applications running on thousands of nodes. However, over time, its capabilities have been extended to handle larger volumes of data and provide real-time analytics. Map-Reduce is a programming model created by Google for processing huge amounts of data using their Big databases.

Map-Reduce is a computing paradigm that facilitates enormous scalability among a large number of servers inside the Hadoop Cluster. Map-Reduce is the soul of Apache Hadoop. The jargon of Map-Reduce indicates two different and distinct tasks on Big data. The first task “Map” converts one set of data into another set of data where the elements are split into tuples known as Key-Value pairs. The second task “Reduce” get the input from the “Map” task i, e. Key-Value pairs and joins these data tuples into a reduced number of tuples. Overall, we get less number of tuples after the Map-Reduce operation on Big data.

3.2 Hierarchical attention networks (HAN)

The Hierarchical Attention network is a deep learning-based classification method. It is built on Bidirectional Recurrent Neural Networks RNNs consisting of GRUs or LSTM. It is built on the Hierarchies with the attention

mechanism where the Upper hierarchies get input from the outputs of the lower hierarchies. There are two concepts (Encoder, Decoder) which are assigned works as Encoder to input information with RNN later and Decoder that connects the encoded contribution to the next recurrent later. In short, HAN consists of a set of Decision Trees, A ordered network(hierarchical), Apriori algorithm and Outlier based multiclass classification. The further explanation is carried out after section 3.2.b.

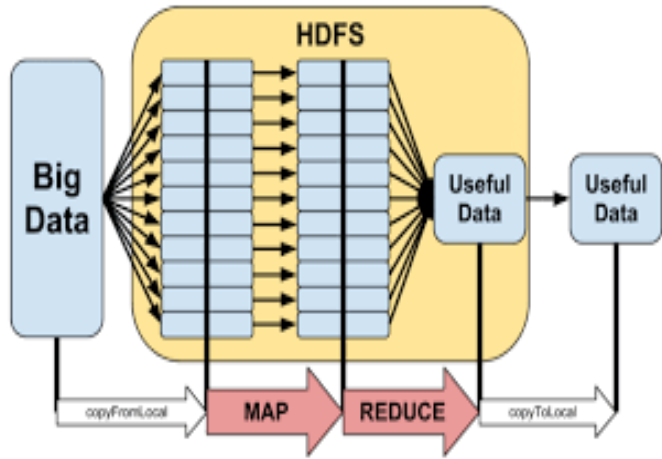


Figure 1. Map Reduce in Hadoop

3.3 Decision trees

Decision trees are one of the widely used supervised classification algorithms in Machine learning and data mining. It takes a set of features as input and generates a Boolean value as output. Decision trees also have the capability of going beyond the binary classification [21, 22].

In Data mining, the Decision tree algorithm adapts the Decision Tree method to generate a learning model, with the Classification rules established on the paths from root node to leaves. The Map-Reduce algorithm as presented in the Figure 1.

$(x, Y)=(x_1,x_2,x_3, \dots , x_n, Y)$. Here $\{x_1,x_2,x_n\}$ are the independent features and Y is the dependent feature. Y holds a binary value, which is also called a label attribute. Decision tree is not sensitive to outliers or noisy data.

This simplicity is the advantage of using this algorithm in our approach in the HAN, while interpreting, the decoder can center on the portions of the encoded input. The Encoder converts the categorical features into numeric vectors second by a couple of LSTM operations. The Encoder and Decoder have different mechanisms. The considered mechanism allows the decoder to ensure only specific parts of the encoder outputs. On the other hand, The Decision trees results are retrieved and utilized by the Hierarchical Consideration Network. This is how a Hierarchical Decision consideration network builds shows in Figure 2. The dataset (set of tables with specific features) is split into training and testing portions. Parallely, the data is cleaned or processed to make the data computable. Some of the examples are treating missing values, imputing the missing values, replacing the text data with numeric data, normalizing the high values, converting categorical data into numeric values, treating the object type values, padding etc. Padding is done when the data has different sizes of sequences (If it is Genomics, it's the Genome sequence [21], if its text classification, it's the length of the text). The padding must be done such that it does not result in loss calculations. The concerned Padding mask containing zeros at its corresponding time steps are the padding valued to be considered. The output is taken as a contribution to the Apriori Algorithm. A brief introduction of the Apriori algorithm is presented in the following section of Figure 3.

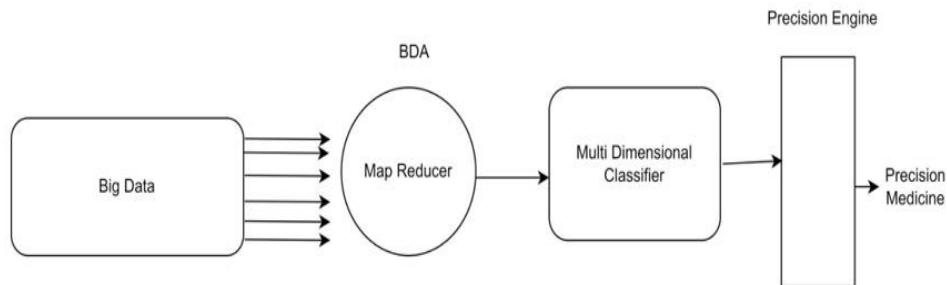


Figure 2. Proposed schema

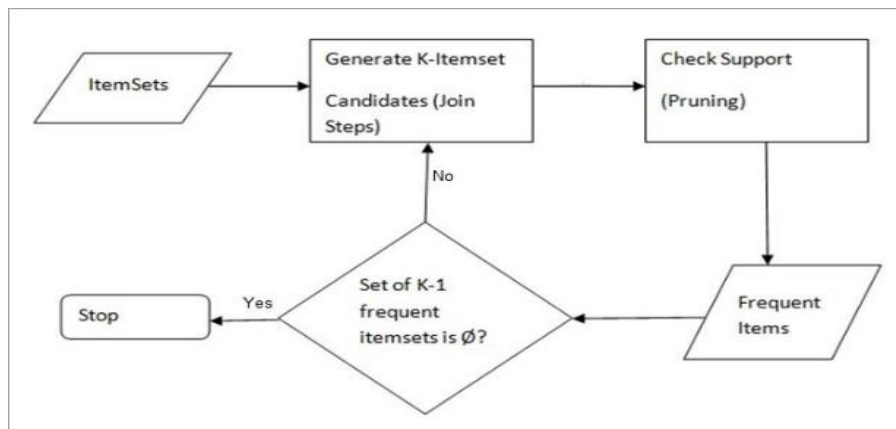


Figure 3. Steps in Apriori algorithm

3.4 Apriori algorithm

Apriori algorithm calculates a flow of steps to identify the most frequent set among the fields in the given dataset. This data mining algorithm follows two steps repeatedly (Join and Prune) till the most frequent set is derived. A base threshold (often called Minimum support) needs to be maintained once the application of this algorithm begins, sometimes it is assumed too. The output of the decision tree attention network algorithm is passed to the Apriori algorithm which Association Rules. The prescribed Association rules use the support and confidence criteria to analyze the given input. This analysis yields the recurrence Associations. In this phase, using the support and confidence, the Apriori algorithm performs Outlier based multi class classification using association rules.

The Support is a value that gives the probability of occurrence of an event. It is detailed below.

$\{x_1 \rightarrow \text{other attributes from } (x_2, x_3, \dots, x_n)\}$.

Support = Number of required Transactions / Total number of transactions in the database. If the support value crosses minimum-support value, then this itemset is added continuously.

The confidence is calculated as below. The Confidence of a rule $P\{x_1 \rightarrow \text{Other attributes from } (x_2, x_3, \dots, x_n)\}$ represents the probability of both antecedents, Resultant for the same transaction. Confidence is calculated with the following formula.

Confidence $\{x_1 \rightarrow \text{other attributes from } (x_2, x_3, \dots, x_n)\}$.

One of the significant applications of Association Rule Mining is to investigate sickness. That relates the alignment to their treatment. The rules derived from Apriori Algorithm are then considered for an ailment to their treatment. Eg: By using Outlier-based Multi class classification, that is explained in the following section.

3.5 Outlier-based multiclass classification

The Result from the Apriori algorithm is given to the Outlier-based Multiclass classification [23, 24]. There are two variants of this phase (i) Multiclass and In-class. The multiclass classification is performed by the results of the Apriori algorithm depending on the prediction through the Support and confidence scores. The dataset is split into Training and Testing portions. Training is performed to detect diabetes in the applicant. The classification of having Diabetes or NO diabetes or Pre-diabetic [24, 25] is performed in this phase. The proposed hierarchical algorithms are expressed as below.

Algorithm

Title: Diabetes classification of Diabetes using HAN

- Step 1. Big data available on the Diabetes data
- Step 2. Load the dataset into Hadoop
- Step 3. Perform Map-Reduce on the given big data
- Step 4. Get the Reduced dataset with a reduced number of features
- Step 5. Initiate Decision Tree Algorithm
- Step 6. Build Attention Network
- Step 7. Run Apriori Algorithm
- Step 8. Produce a Set of Rules
- Step 9. Perform Outlier based Multi-class classification
- Step 10. Perform Diabetes Classification
- Step 11. If diagnosed (Diabetes or Pre-Diabetic) go to Step 12

Step 12. Initiate Intelligent Precision Engine

Step 13. Else: No Prescription to the user

Step 14. End

3.5.1 The intelligent precision medicine engine

The second phase of our work deals with the precision medicine engine [25, 26]. Once diabetes is diagnosed in Phase one of this work, it's time to provide precision medicine [27-29].

The diabetes medicine will be prescribed based on the level of diabetes.

As per the E-Medicine information from Medscape, there are

- "Type 1" Diabetes Mellitus
- "Type 2" Diabetes Mellitus (Often called as Pre-Diabetes)
- Pediatric Diabetes Mellitus
- Gestational Diabetes Mellitus

In the computer science perspective we treat each type as a class. Thus we adapt Multi-Class classification. The proposed component Intelligent precision medicine for Diabetes implemented with advanced recommender algorithms which can mention the micrograms of the medicine (Glycomet) or Insulin. More work on this component is part of our future work. The following section explains the Experimentation details.

4. EXPERIMENTATION AND DISCUSSION

We have utilized a Big data set with more than one lakh medical records with more than 55 attributes. As we know that the size of medical data is growing exponentially, working on a huge dataset for research is a good contribution.

Dataset description

The Diabetes dataset considered in this experiment is taken from the study [3]. It has 101766 records with the following fields.

```
['encounter_id',  
'patient_nbr',  
'race',  
'gender',  
'age',  
'weight',  
'admission_type-id',  
'discharge_disposition_id',  
'admission_source-id',  
'time_in_hospital',  
'payer_code',  
'medical_specialty',  
'num_lab_procedures',  
'num_medications',  
'number-outpatient',  
'number_inpatient',  
'diag-1',  
'diag-2',  
'diag-3',  
'number_diagnoses',  
'max_glu_serum',  
'A1cresult',  
'metformin',
```

'repaglinide',
'nateglinide',
'chlorpropamide',
'glimepiride',
'acetohexamide',
'glipizide',
'glyburide',
'tolbutamide',
'pioglitazone',
'rosiglitazone',
'acarbose',
'miglitol',
'troglitazone',
'tolazamide',
'examide',
'citoglipton',
'insulin',
'glyburide-metformin',
'glipizide-metformin',
'change',
'diabetesMed',
'readmitted']

Out of the above 50 attributes, we may not need all features. Also, it will take high computational power to perform any analysis. The label attribute is “DiabetesMed” which is used for Classification. It indicates the existence of diabetes. The last feature is “Readmitted” which does not give any

contribution to our work in this paper. The attributes with Metformin indicate the medicine, we can also see other medicine Insulin. There are some missing values in some fields, thus we have dropped those entries are mentioned in Table 1.

- The Map-Reduce on Hadoop framework on this dataset produced a record set of 18566 entries. The Map-Reduce on Hadoop has reduced the dataset to a decent extent. The number of attributes as well as the dataset size is also optimized [30]. The Map-Reduce on Hadoop availed 14 attributes.

- They are age, Gender, Weight, 'max_glu_serum', 'metformin', 'repaglinide', 'chlorpropamide', 'citoglipton', 'insulin', 'glyburide-metformin', 'glipizide-metformin', 'metformin-rosiglitazone', 'metformin-pioglitazone', 'diabetesMed'.

- The Resulted Reduced data is experimented with HAN.

- The Apriori algorithm yielded the implications that “Insulin” and “diabetesMed” are associated. The resultant Support was 0.66 and Confidence was 1.0.

- Also, we have experimented with general Machine learning classification algorithms the Performance results are as follows.

- The Decision Tree algorithm formulated the derivation that {Insulin, Age} being implied to {diabetesMed}.

Table 1. The performance measures are tabulated

Algorithms /Performance measures	Proposed Schema of HAN-based Classification	XGboost	Random Forest Classifiers
Precision	0.98	0.907	0.79
Recall	0.98	0.9241	0.79
F1 Score	0.98	0.924	0.78
Accuracy_Score	0.98	0.913	0.79

The results are depicted in the below Figure 4.



Figure 4. Performance comparison

We can see that the Traditional Machine learning classifier Random Forest has performed less compared to other methods. It is due to the higher information retrieval capacity of the proposed HAN method. HAN performed better because these networks show a stable performance by protecting against outliers. The XGBoost also performed decently, because of the large dataset. If the data set is smaller, the performance might be much less with XGBoost and Random forest classifier.

5. CONCLUSION AND FUTURE WORKS

In this paper summarizes a quantitative measure can be done on any framework for precision medicine. It can be added as a Future Research objective. The second phase of our work i.e. “building intelligent precision medicine Engine” can be taken as an upcoming research problem, which has a wide scope of acceptability. For each disease, there will be a unique precision medicine engine. We continue our research by building this component and performing a quantitative assessment.

REFERENCES

- [1] MacEachern, S.J., Forkert, N.D. (2021). Machine learning for precision medicine. *Genome*, 64(4): 416-425. <https://doi.org/10.1139/gen-2020-0131>
- [2] Dua, D., Graff, C. (2019). UCI Machine learning repository. Irvine, CA: University of California, School of Information and Computer Science. <http://archive.ics.uci.edu/ml>.
- [3] 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. <https://doi.org/10.3390/s18072183>

- [4] Gupta, B., Kumar, A., Dwivedi, R.K. (2018). Big data and its applications-a review. In 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). IEEE, pp. 146-149. <https://doi.org/10.1109/ICACCCN.2018.8748743>
- [5] Gao, P.L., Han, Z.M., Wan, F.C. (2020). Big data processing and application research. In 2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM), Manchester, pp. 125-128. <https://doi.org/10.1109/AIAM50918.2020.00031>
- [6] Rao, T.R., Mitra, P., Bhatt, R., Goswami, A. (2019). The big data system, components, tools, and technologies: A survey. *Knowledge and Information Systems*, 60: 1165-1245. <https://doi.org/10.1007/s10115-018-1248-0>
- [7] Bochicchio, M., Cuzzocrea, A., Vaira, L. (2016). A big data analytics framework for supporting multidimensional mining over big healthcare data. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 508-513. <https://doi.org/10.1109/ICMLA.2016.0090>
- [8] Bochicchio, M.A., Cuzzocrea, A., Vaira, L., Longo, A., Zappatore, M. (2018). Multidimensional mining over big healthcare data: a big data analytics framework. In SEBD.
- [9] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E. (2016). Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1480-1489.
- [10] Anil, G.R., Moiz, S.A. (2020). Blockchain enabled smart learning environment framework. In Advances in Decision Sciences, Image Processing, Security and Computer Vision: International Conference on Emerging Trends in Engineering (ICETE). Springer International Publishing, 2: 728-740. https://doi.org/10.1007/978-3-030-24318-0_83
- [11] Solis-Herrera, C., Triplitt, C., Reasner C., DeFronzo, R.A., Cersosimo, E. (2018). Classification of Diabetes Mellitus. In: Feingold KR, Anawalt B, Boyce A, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. <https://www.ncbi.nlm.nih.gov/books/NBK279119/>.
- [12] Agarwal, A., Pritchard, D., Gullett, L., Amanti, K.G., Gustavsen, G. (2021). A quantitative framework for measuring personalized medicine integration into us healthcare delivery organizations. *Journal of Personalized Medicine*, 11(3): 196. <https://doi.org/10.3390/jpm11030196>
- [13] American Diabetes Association. (2018). 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. *Diabetes Care*, 41(Supplement_1): S13-S27. <https://doi.org/10.2337/dc18-S002>
- [14] American Diabetes Association. (2011). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 34(Supplement_1): S62-S69. <https://doi.org/10.2337/dc10-S062>, accessed on Jan. 12, 2023.
- [15] Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1): 1-25. <https://doi.org/10.1186/s40537-019-0217-0>
- [16] Gou, X., Xu, Z. (2021). An overview of big data in healthcare: multiple angle analyses. *Journal of Smart Environments and Green Computing*, 1(3): 131-145. <https://doi.org/10.20517/jsegc.2021.07>
- [17] Haleem, A., Javaid, M., Singh, R.P., Suman, R., Rab, S. (2021). Blockchain technology applications in healthcare: An overview. *International Journal of Intelligent Networks*, 2: 130-139. <https://doi.org/10.1016/j.ijin.2021.09.005>
- [18] Hashim, F., Harous, S. (2021). Precision medicine blockchained: A review. In 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM). IEEE, pp. 48-52. <https://doi.org/10.1109/ICCAKM50778.2021.9357760>
- [19] Beam, A.L., Kohane, I.S. (2018). Big data and machine learning in health care. *Jama*, 319(13): 1317-1318. <https://doi.org/10.1001/jama.2017.18391>
- [20] De Silva, D., Burstein, F., Jelinek, H.F., Stranieri, A. (2015). Addressing the complexities of big data analytics in healthcare: the diabetes screening case. *Australasian Journal of Information Systems*, 19. <https://doi.org/10.3127/ajis.v19i0.1183>
- [21] Yang, F.J. (2019). An extended idea about decision trees. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, pp. 349-354. <https://doi.org/10.1109/CSCI49370.2019.00068>
- [22] Domathoti, B., Ch, C., Madala, S.R., Berhanu, A.A., Narasimha Rao, Y. (2022). Simulation analysis of 4G/5G OFDM systems by optimal wavelets with BPSK modulator. *Journal of Sensors*, 2022: 8070428. <https://doi.org/10.1155/2022/8070428>
- [23] Prospero, M., Min, J.S., Bian, J., Modave, F. (2018). Big data hurdles in precision medicine and precision public health. *BMC Medical Informatics and Decision Making*, 18: 1-15. <https://doi.org/10.1186/s12911-018-0719-2>
- [24] Ndirangu, D., Mwangi, W., Nderu, L. (2019). An ensemble model for multiclass classification and outlier detection method in data mining. *Journal of Information Engineering and Applications*, 9: 22-30. <https://doi.org/10.7176/JIEA>
- [25] König, I.R., Fuchs, O., Hansen, G., von Mutius, E., Kopp, M.V. (2017). What is precision medicine? *European Respiratory Journal*, 50(4). <https://doi.org/10.1183/13993003.00391-2017>
- [26] Skyler, J.S., Bakris, G.L., Bonifacio, E., Darsow, T., Eckel, R.H., Groop, L., Groop, P.H., Handelsman, Y., Insel, R.A., Mathieu, C., McElvaine, A.T., Palmer, J.P., Pugliese, A., Schatz, D.A., Sosenko, J.M., Wilding, J.P.H., Ratner, R.E. (2017). Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes*, 66(2): 241-255. <https://doi.org/10.2337/db16-0806>
- [27] Wu, H., Tremaroli, V., Schmidt, C., Lundqvist, A., Olsson, L. M., Krämer, M., Gummesson, A., Perkins, R., Bergström, G., Bäckhed, F. (2020). The gut microbiota in prediabetes and diabetes: A population-based cross-sectional study. *Cell Metabolism*, 32(3): 379-390. <https://doi.org/10.1016/j.cmet.2020.06.011>
- [28] Landmark Precision Medicine Research Effort Releases its First Genomic Dataset (ucsd.edu). <https://health.ucsd.edu/news/press-releases/2022-03-17-landmark-precision-medicine-research-effort-releases-its-first-genomic-dataset/>, accessed on Dec. 12, 2022.
- [29] Wang, L., Alexander, C.A. (2019). Big data analytics in healthcare systems. *International Journal of*

Mathematical, Engineering and Management Sciences,
4(1): 17. <https://doi.org/10.33889/IJMEMS.2019.4.1-002>

[30] Khardori, R., Griffing, G. (2017). Type 1 diabetes mellitus treatment & management. Available: Type, 2. <https://emedicine.medscape.com/article/117739-treatment>.