

Tailored Medicine for Patients based on Data Mining and Machine Learning Techniques

Jeffrey Giftson Jeshurun¹, Nallarasan V², Dakshan Cumar³

¹Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur Campus

²Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur Campus

³Department of Information Technology, SRM Institute of Science and Technology, Kattankulathur Campus

ABSTRACT

Healthcare sector keeps on improvising with innovative technology day wise, as everyday possibility of new diseases is emerging and biomedical automation and data analysis is needful to reduce the human efforts. In traditional approach, the same drugs, the same tablets and the same medical treatments is being provided to every patient with the same disease. This approach is not a successful one as every human system would be different from others based on their individual routine life activities, family inheritance, genetics, RNA sequence factors, food habitats, working nature etc. Thus same drug would work effectively in certain human body and same don't have any effect in certain human body. Thus personalized medicine recommendation is needed in analyzing the human traits. This proposed system helps the doctor to provide right medicines to right patients based on each their trait analysis. Thus this proposed system would create an evolution in the healthcare service sector and decrease the mortality rates providing good, healthy life.

Keywords

Human DNA: gene, proteins, tailored medicine, lifestyle, family inheritance, RNA sequence

Introduction

The personalized medicinal treatment is nothing but suggestion of specific medicines to individuals who belong to a group of same genetic parameters, food habitats, working nature. While suggestion of medicines the doctors have to analyze the secondary parameters like impact of the disease, disease stage and patient body stability. Usually in past people state "the doctors while doing treatment they got to see the type of patient not type of disease" because the treatment would be effective and success if the doctors recommend right drugs to the right patients. Thus explaining the effective approach for the physicians that recommendation of right drugs not only analyzing the disease name but exploring primary factors like family inheritance, gene report, lifestyle, working nature and RNA sequences [1]. Exploring the above human traits will provide completed knowledge of the patient to the doctor, hence the doctor would get a good knowledge on the drug name and dosage which the respective patient body can able to adapt producing good effects [2].

Due to increasing patient numbers day to day, new medicine recommendation system is needful invoking genetic based inspection. Emergence of large volume of health data, technology like data analytics make the health sector research open many dimensions. Researches state mining of human DNA, medicinal history, family disease inheritance, working nature, routine activities like physical activities might be helpful in analyzing and suggestion of tailored medicines for patients for the same disease [3]. In the initial training phase post drug reaction is also analyzed for

identifying the effectiveness of the drug in human bodies. Few research articles state that the concept of the tailored medicine for patient analyzing their medicinal and personal traits would be able to have great outcome for deadly diseases like increased sugar level, cancer, cardiac diseases etc. Because of the emergence of Internet, the health data availability becomes simple and feasible. The challenge in this is to mine the huge dataset to discover the pattern associated to identify the specific medicine for the tailored medicine [4-5]. Manual based processing of huge data would be difficult, time consuming, lead to many human errors. Human errors in health sector would lead to a sensitive issue. Thus instead of manual based processing, data mining based patient personal, medicinal traits analysis and tailored medicine suggestion has a greater scope to reduce the mortality rate [6].

Analyzing different types of personal and medical traits of huge sample of different patients involves huge size of data, increased efforts, data analytics experts to process the data and identify the hidden relationship patterns between human traits and medicines. Also for analysis we have to obtain data of different set of ages to discover the right medicine for group of individuals with same human traits and also study the reaction of the drug in individuals [7].

Thus despite of the challenge of processing huge set of health data of different ages, the technology emergence and adaption of technology in the health sector makes the data availability easy and emergence of data analytics, machine learning algorithms make data analysis feasible for effective decision making within a short period of time [8]. Thus the proposed tailored medicine recommendation would be able to identify right drug for the right patients reducing the mortality rate. In this proposed project, tailored medicine based recommendation is performed for diabetes disease detection which would be helpful to control the sugar level at the early stage.

Related Works

This article briefs about how Machine Learning, which is a division of Artificial Intelligence, is helpful in analyzing diagnostic test results and other resources to generate personalized medicine/treatment. Some challenges with Artificial Intelligence in healthcare mentioned in this review include explainability, liability and privacy. In order to address the explainability challenge, we have to develop models which can explain "why" the particular medicine or treatment has been recommended for the particular individual. Liability and privacy issues can be solved by using technologies and regulatory bodies [9]. This article briefs specific medicine recommendation for the patients based on grouping the patients with regard to type of diseases, age, risk of treatments, body response level to the specific medicines. This paper analyzes the various biomarkers with regard to lung cancer within short time [10]. This paper used real world database to classify patients with Cox-2 inhibitors i.e., VioxxTM and CelebrexTM. The authors applied various statistical models for analyzing medicines to provide to these patients who are treated as critical stage of getting heart attack [11]. This paper performed an important research on the survival rate of breast cancer patients by collection and analysis of their historical data's. The problem which they addressed was the data collected was not labelled types, hence the researchers spend 5 years for labelling the data analyzing the historical inputs into survived and not survived. The real time health data are collected by signing confidential documents with the medical labs, doctors, patients. For training the model to identify the patterns they have used semi-supervised learning (SSL) [12-13].

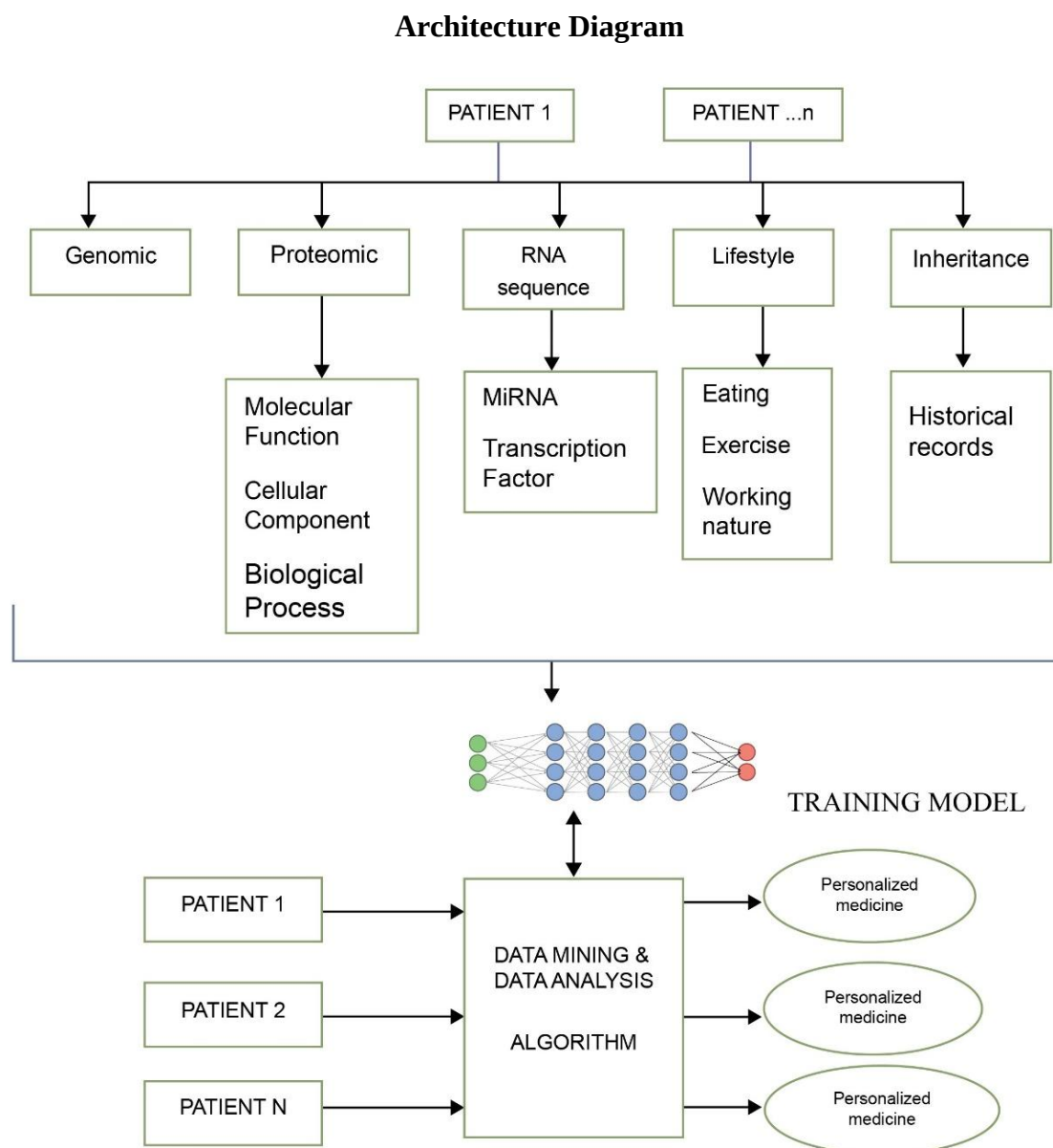


Figure 1. briefs our proposed architecture in the analysis and suggestion of personalized medicine for the patients. The analysis stage is given much importance invoking genetic aspects, inheritance, lifestyle, working nature and food habits.

Proposed Methodology

Due to the emergence of new diseases and increasing volume of patients, tailored medicine would be a best practice for disease prevention and reduce death rates. Many previous research explains that the tailored medicine recommendation by deep analyzing of patient's inheritance, life style, working nature, gene and RNA analysis helps in prevention and enhancement of critical diseases such as cancer, diabetics, cardiac diseases etc. In our proposed methodology, gene analysis is an important factor to predict the human body enzymes and drug reaction level in the respective patient body. In our application, the gene and protein analysis is performed by patients entering their genetic test details. The human proteomic has three factors specifically

cellular component (CC), biological process (BP) and molecular function (MF) [14]. The human genomic is the regulatory module in the human body involving miRNA, transcription factor and gene. Based on the respective human genetic input data, extrinsic and intrinsic enzymes are defined. After the human gene and protein analysis, the patients low and high level of enzymes are identified. Based on this the patient risk factor is identified. The patients are differentiated into different level of groups by taking into account of certain parameters like stage of disease, other diseases present, risk, drug reaction level etc. [15]. The adverse drug reactions invoke machine learning technique to train and identify the best healing medicine for a respective group of patients. Thus recommendation of personalized medicine can be achieved with promising results through machine learning technique [16]. In the proposed system, Iterative Multiplicative Updating Algorithm is implemented to check the optimization function between the regulatory modules. For training the module, we have used fuzzy logic to identity the best drug, vaccine for a respective human, based on the genetic, lifestyle, work nature, food habitats, inheritance. MOAL (Multi Ontology Data Mining at all Level) algorithm is been proposed to identify the extrinsic, intrinsic problem of the human protein bodies between molecular function (MF), biological process (BP) and cellular component (CC). In real time, training module will take much time as because of the depth in the subject and this proposed system cold crate a revolution in the health sector with reducing mortality rates.

Implementation Results

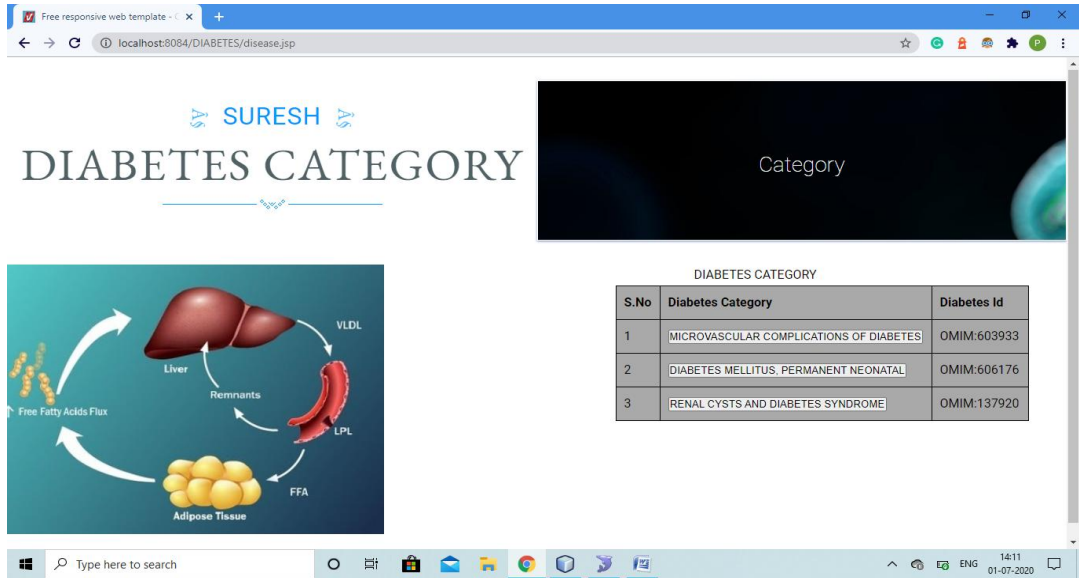
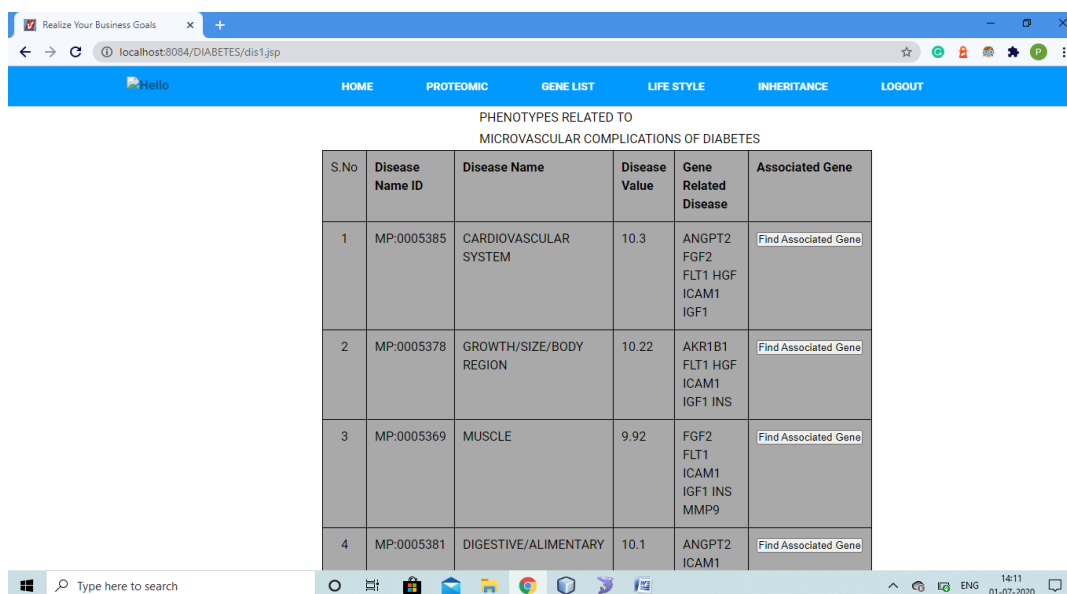


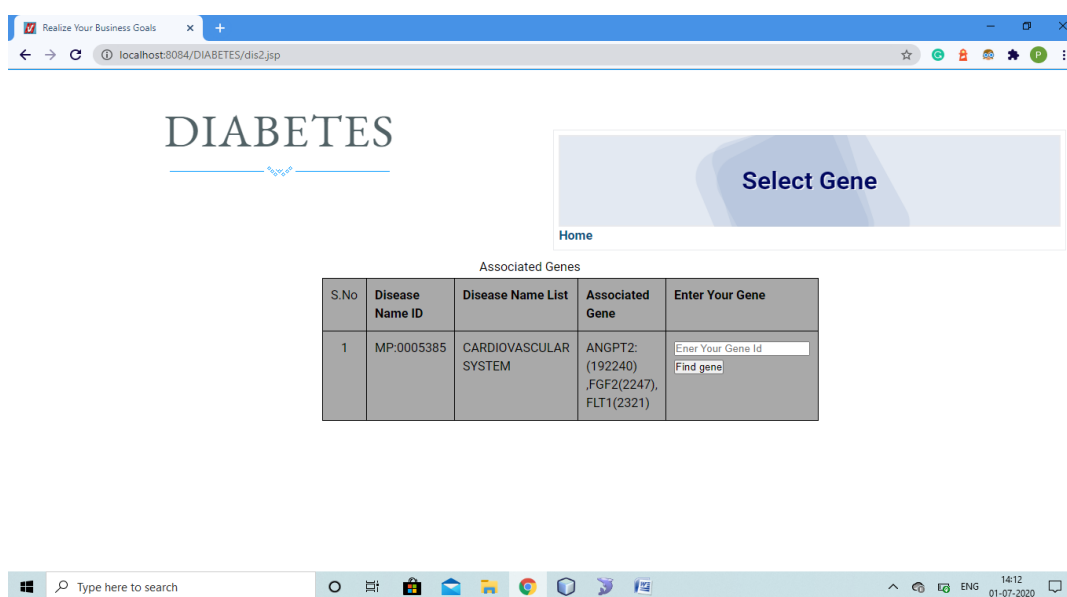
Figure 2. Disease names



PHENOTYPES RELATED TO
MICROVASCULAR COMPLICATIONS OF DIABETES

S.No	Disease Name ID	Disease Name	Disease Value	Gene Related Disease	Associated Gene
1	MP-0005385	CARDIOVASCULAR SYSTEM	10.3	ANGPT2 FGF2 FLT1 HGF ICAM1 IGF1	Find Associated Gene
2	MP-0005378	GROWTH/SIZE/BODY REGION	10.22	AKR1B1 FLT1 HGF ICAM1 IGF1 INS	Find Associated Gene
3	MP-0005369	MUSCLE	9.92	FGF2 FLT1 ICAM1 IGF1 INS MMP9	Find Associated Gene
4	MP-0005381	DIGESTIVE/ALIMENTARY	10.1	ANGPT2 ICAM1	Find Associated Gene

Figure 3. Gene related to disease



DIABETES

[Home](#)

Select Gene

Associated Genes

S.No	Disease Name ID	Disease Name List	Associated Gene	Enter Your Gene
1	MP-0005385	CARDIOVASCULAR SYSTEM	ANGPT2: (192240), FGF2(2247), FLT1(2321)	Enter Your Gene Id Find gene

Figure 4. User providing his gene report data

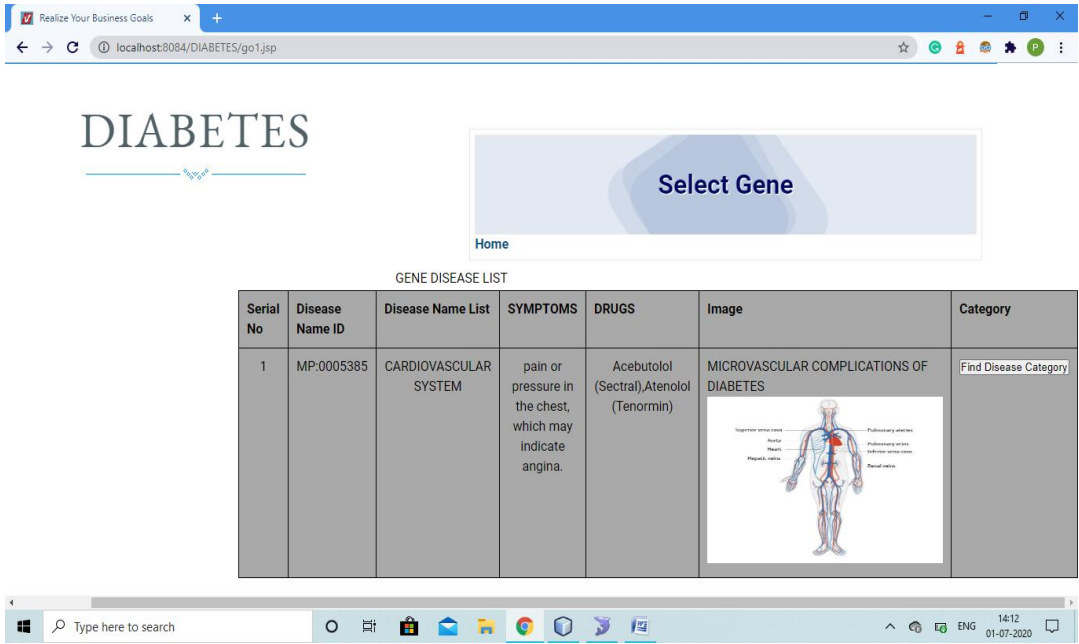


Figure 5. Disease name with symptoms and drug

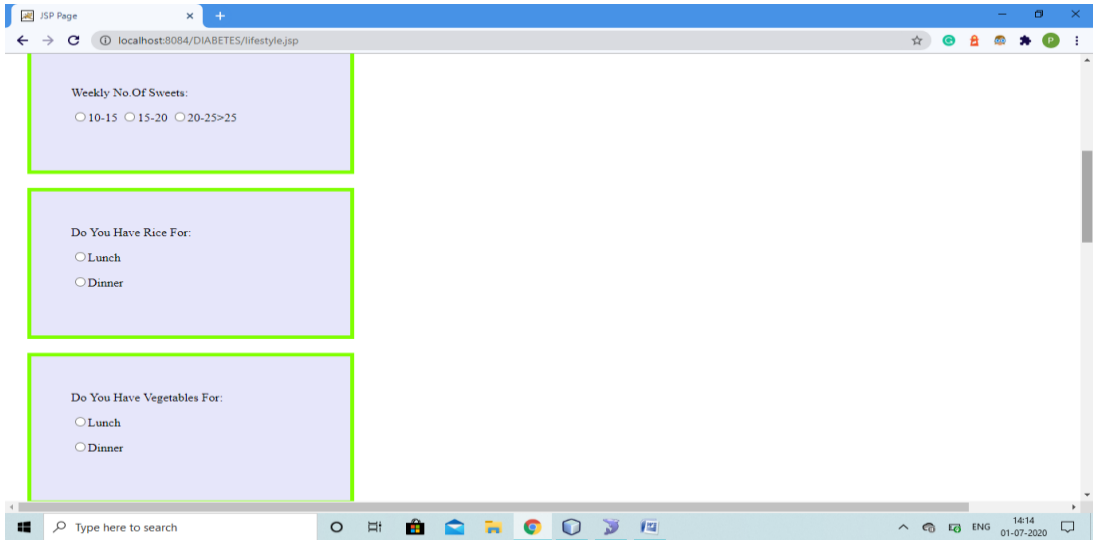


Figure 6. User lifestyle input

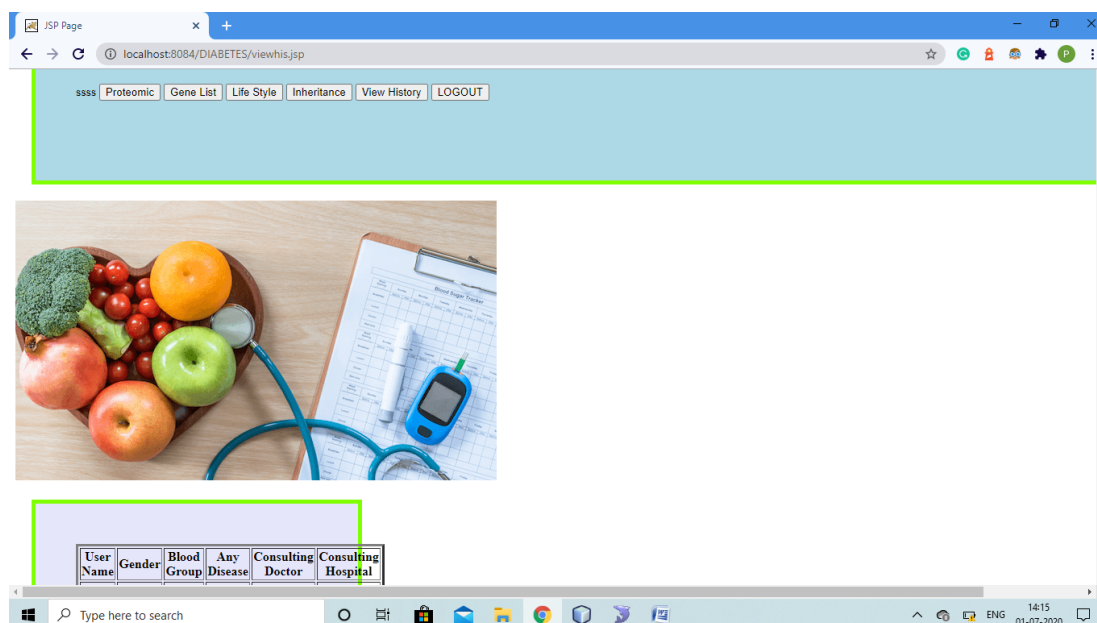


Figure 7. Personalized medicine is recommended

Table 1. Gene and Protein data

Gene id	MiRNA	TF	Gene	Weighted Confidence	Cross ontology category
64324	MI0000060	chr11	NSD1	0.50	BP-CC-MF
1028	MI0000061	chr22	CDKN1C	0.35	BP-CC-MF
105259599	MI0000062	chr22	H19-ICR	0.74	BP-CC-MF
100506658	MI0000064	chr21	OCLN	2.00	BP-CC-MF
55630	MI0000065	chr9	SLC39A4	0.64	BP-CC-MF
1130	MI0000066	chr19	LYST	0.75	BP-CC-MF
2517	MI0000067	chr9	FUCA1	0.89	BP-CC-MF
2629	MI0000068	chrX	GBA	0.67	BP-CC-MF
4688	MI0000263	chr12	NCF2	0.84	BP-CC-MF
2720	MI0000265	chr19	GLB1	0.98	BP-CC-MF

Table 2. Performance evaluation of the proposed system

Sample Inputs	Execution Time (s)
BECKWITH-WIEDEMANN SYNDROME	4
BILE ACID SYNTHESIS DEFECT, CONGENITAL, 3	3.99
LYSINURIC PROTEIN INTOLERANCE	5.20
MAST CELL DISEASE	4.25
Spinal Muscular Atrophy-1 (SMA1)	4.75

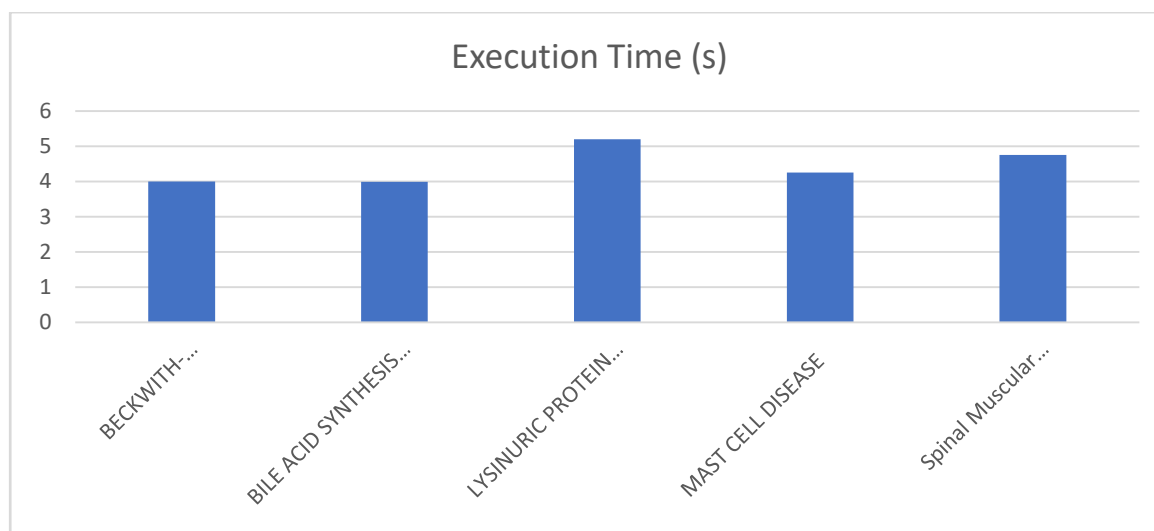


Figure 8. Execution performance evaluation of the proposed system

Conclusion

Personalized medicine has a huge scope in the health sector which can help doctors to provide right diagnosis to the diseases at the early stage analyzing each patient's body. The proposed methodology recommends medicines for patients based on age, weight, inheritance, genomic, proteomic, RNA sequences, life style, inheritance, food habitats, risk factor etc. Applying machine learning techniques provides promising results reducing the death rate and saving human lives.

References

- [1] A. J. Vargas and C. C. Harris. (2016). Biomarker development in the precision medicine era: lung cancer as a case study. *Nature Reviews Cancer*, 16(8), p. 525.
- [2] Wei, Z., Wang, K., Qu, H. Q., Zhang, H., Bradfield, J., Kim, C., Frackleton, E., Hou, C., Glessner, J. T., Chiavacci, R., Stanley, C., Monos, D., Grant, S. F. A., Polychronakos, C., & Hakonarson, H. (2009). From disease association to risk assessment: An optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genetics*, 5(10), [e1000678].
- [3] J. Kim and H. Shin. (2013). Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data. *Journal of the American Medical Informatics Association*, 20(4), 613–618
- [4] E.M.Antman and J.Loscalzo. (2016). Precision medicine in cardiology. *Nature Reviews Cardiology*, 13(10), p. 591.
- [5] D. Bzdok and A. Meyer-Lindenberg. (2017). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.
- [6] J. Davis, E. Lantz, D. Page, J. Struyf, P. Peissig, H. Vidaillet, and M. Caldwell. (2008). Machine learning for personalized medicine: Will this drug give me a heart attack. *in the Proceedings of International Conference on Machine Learning (ICML)*.

- [7] P. Kieseberg, H. Hobel, S. Schrittwieser, E. Weippl, and A. Holzinger.(2014). Protecting anonymity in data-driven biomedical science. In: *Holzinger A., Jurisica I. (eds) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. Springer.* 301–316.
- [8] Y. LeCun, Y. Bengio, and G. Hinton. (2015).Deep learning.*nature*, 521(7553), p. 436.
- [9] K. Paranjape, M. Schinkel and P. Nanayakkara.(2020). Short Keynote Paper: Mainstreaming Personalized Healthcare–Transforming Healthcare Through New Era of ArtificialIntelligence," in *IEEE Journal of Biomedical and Health Informatics*,24(7), 1860-1863.
- [10] X. Glorot and Y. Bengio.(2010). Understanding the difficulty of training deep feedforward neural networks. in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics.* 249–256.
- [11] P. Mamoshina, A. Vieira, E. Putin, and A. Zhavoronkov.(2016). Applications of deep learning in biomedicine.*Molecular pharmaceutics*, 13(5), 1445–1454.
- [12] E. Gawehn, J. A. Hiss, and G. Schneider.(2016). Deep learning in drug discovery.*Molecular informatics*, 35(1), 3–14.
- [13] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke. (2018). The rise of deep learning in drug discovery. *Drug discovery today*.
- [14] A. Gottlieb, G. Y. Stein, E. Ruppín, and R. Sharan. (2011). Predict: a method for inferring novel drug indications with application to personalized medicine.*Molecular systems biology*, 7(1), p. 496.
- [15] S. Ekins. (2016). The next era: Deep learning in pharmaceutical research.*Pharmaceutical research*,33(11), 2594–2603.
- [16] M. H. Segler and M. P. Waller. (2017). Neural-symbolic machine learning for retrosynthesis and reaction prediction.*Chemistry-A European Journal*, 23(25), 5966–5971.
- [17] P. Vijayakumar, V. Nallarasán and A. Joshua Jafferson. (2019). Fuzzy Logic based Reliable Spectrum Sensing-A SDR Implementation. *Journal of Advanced Research in Dynamical and Control Systems*, 11(11), 137-142.
- [18] V.Nallarasán,K.Kottilingam. (2020). Enhanced security in IoT Networks using ensemble methods in cognitive radio approach. *International Journal of Emerging Trends in Engineering Research*, 8(8).
- [19] Anand, L., Nallarasán, V., Krishnan, M.B.M., Jeeva, S. (2020). Driver profiling-based anti-theft system. *AIP Conference Proceedings*, 2282, 020042.
- [20] Prabakaran, J., Damodharan, D., Anandhan, K., Suresh Kumar, N., Nallarasán, V. (2020). Enhanced predictive learning approaches for tweaked diet proposal system in health care. *International Journal of Emerging Trends in Engineering Research*, 8(9), 6334–6341, 229.
- [21] V Nallarasán, A Shobanadevi, J Prabakaran, S Prabakaran. (2017). A Distributed Cluster Algorithm for Cluster-head failures in UWSNs. *International Journal of Pure and Applied Mathematics*, 117 (15), 1121-1127.