European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 2020

TAILORED MEDICINE FOR DISEASED PATIENT BASED ON DATA MINING AND MACHINE LEARNING TECHNIQUES

Jaichandran R, Shanmuganathan K.L, Leelavathy S, Sidharth Baburaj, Murshid Iban and Rehil Rej

Department of Computer Science and Engineering, Aarupadai Veedu Institute of Technology,

Vinyaka Missions Research foundation (Deemed to be University), Paiyanoor, Tamil Nadu , India

rjaichandran@gmail.com, shanmuganathankl@avit.ac.in, leelavathy@avit.ac.in, sidharthnambiar23@gmail.com

ABSTRACT:

Healthcare sector is growing with new technology day to day as new diseases are emerging and improvisation is demandful. In the current medicine recommendation, same vaccine, drugs are suggested to all the patients with disease symptoms and physician doesn't analyze the patient's gene, lifestyle and inheritance of each patients. In this paper, we have proposed recommendation of personalized drugs, medicine to patients by analyzing patients DNA – genomic, proteomic, inheritance, lifestyle and RNA sequences so that tailored medicine can be suggested to each patients thus providing right medicine for right patients. Thus suggestion of personalized medicine to patients would improvise healthcare services and reduce death rates among human life's.

Keywords: DNA – genomic, proteomic, inheritance, lifestyle and RNA sequences, personalized medicines.

1. INTRODUCTION

Personalized medicine is also termed as precision medicine, which is nothing but tailored medicines, drugs, vaccines to specific group of victims based on their risk, disease, disease stages. There is old quote that says, it is important to analyze the type of person the disease has rather than the type of disease the person has. Thus this quote clearly states that personalized tailored medicine should be the focus for the doctor while recommending the drugs, this invokes genetics, inheritance, lifestyle, etc., rather than checking only on the disease phenotypes.

With the rapid growth of increasing patients data, genetic analysis technology emergence in health care sector is making personalized drug recommendation for patients viable leaving the traditional approach. The personalized medicine has the potential to address complex disease by deep analysis of genetical data, medical history, inheritance, lifestyle etc [1]. Many researches say this proposed system can address complex diseases such as diabetics, cancer, heart diseases, psychiatric diseases etc. The main challenge in the proposed system is that predicting specific vaccines, medicines to individuals needs to mine huge set of data which is complex to perform manually [2]. Thus data mining based software is demandful to analyse the patient's data and recommend tailored medicines for the specific group of patients. Example, the proposed software has to analyze many parameters like blood pressure, food habits, sugar level, working habitat while recommending drugs for a cardiovascular disease. Thus when this proposed system applied in real time, this invokes deep data patterns, hidden relationships which need lot of efforts and expertise. Example, Recommendation of personalized medicine needs huge set of data from age, weight, blood pressure, medical history, genomic data, working nature, inheritance etc.. thus the data involved are not in large in size there are dimensionally varied, unstructured [3].

Thus with the recent advancements in heath sector, patients data are available vast which can be used for analytics to recommend personalized medicine. Thus recommendation of tailored medicine for patient's

analyzing the related parameters would be an evolution in health care services providing reduction rate of deaths, able to identify dangerous diseases at the early stage itself.

2. RELATED WORKS

[4] This research paper, explains about precision medicine's for patients based on classifying the patients based on the disease type, risk level, response rate to the treatments. Though omics based technique able to identify the biomarks faster but the reproducibility rate is poor. This research paper also summarize different type of biomarks with regard to lung cancer as an example.

[5] This research paper research about breast cancer survivability of patients with breast cancer using machine learning techniques based on the historical patients data. However the data available are not labelled ones, hence these authors took 5 years to label the data for each patients like survived or not survived. Thus the authors were able to obtain these labeled data with greater efforts by maintaining confidential agreement with doctors, clinics and patients to preserve their patients privacy. They used semi-supervised learning (SSL), a recent machine learning algorithm for data analytics and recommendation.[7]

[6] This paper majorly focus on real world database invoing many statistical models to identify patients with Cox-2 inhibitors, such as VioxxTM and CelebrexTM. These patients are considered as high risk for heart attack.

3. METHODOLOGY

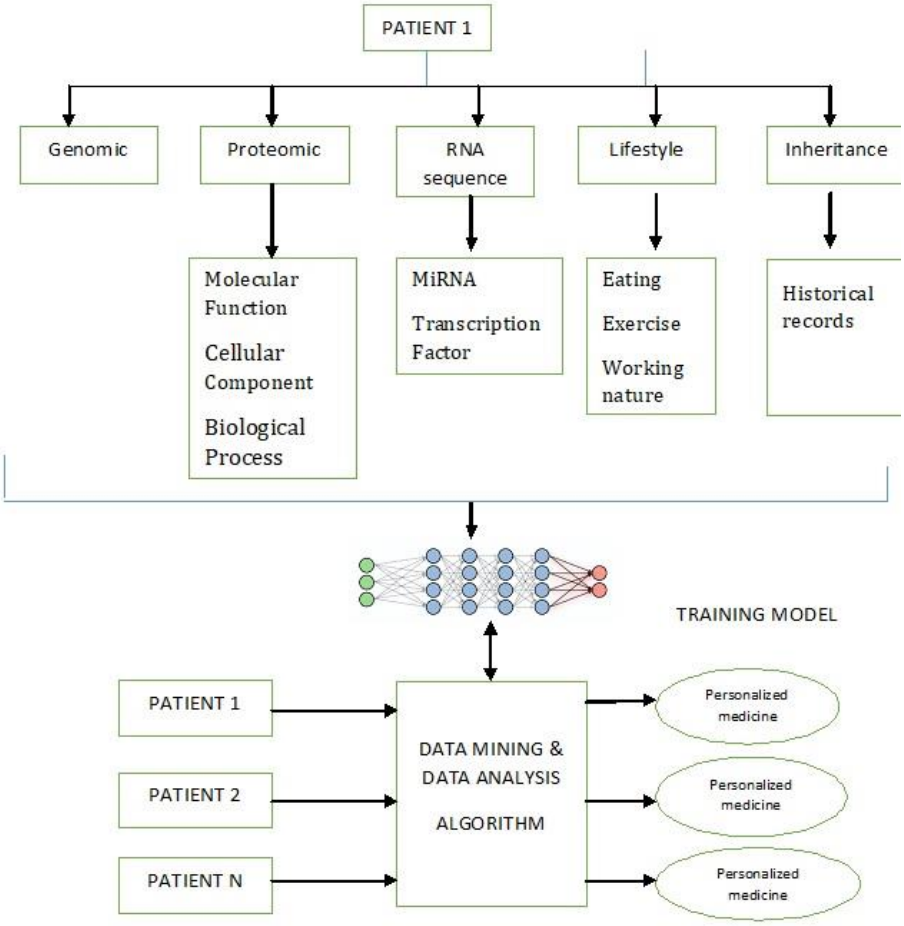

Fig 1. Methodology

Fig 1 explains our proposed architecture of recommending personalized medicne for patients based on analysint their gentical aspects, inheritance, lifestyle, working nature and food habitats.

With the rapid growing size of medical database personalized medicine for patient can be an alternative to the traditional medication. Personalized medicine has wide applications in recommending drugs, vaccines,

medicines, treatments for complex diseases like heart disease, diabetics, cancer etc. Our proposed system, recommends tailored medicine to patients effectively analyzing the DNA – genomic, proteomic, inheritance, lifestyle and RNA sequences may be an evolution in medicine recommendation saving human life's.

3.1. PATIENT DETAILS DATA

In this module the patient has to provide personal, medical data like age, weight, blood pressure, inheritance details, food habitats, working nature, lifestyle, genomic and proteomic data, RNA sequences etc (gene test details) etc. Proteomic data includes *cellular component, molecular function* and *biological process* values of the patients analyzing which enzyme is intrinsic and extrinsic. The patient Co-Regulatory modules between miRNA (microRNA), TF (Transcription Factor) and gene on function level with multiple genomic data.

3.2.IDENTIFICATION OF RISK FACTOR

In this module we cluster the patient into different groups based on the disease, stage of disease, risk, respond rate to treatment. On basic level of clustering the patients would be undergoing another level of clustering in which the patients are again grouped based on the drug reactions, thus adverse drug reaction are collected on a regular basis and based on that patterns machine learning technique is applied to identify best medicine for that specific group of patient's.

3.3.DATA MINING AND DATA ANALYSIS ALGORITHM

In this module we analyze the genomic and proteomic data analysis for patients using **Collaborative filtering** algorithm. This mainly identify the disease based on protein values, biological process values, molecular function values, cellular component values. Based on the BP, CC and MF values the Intrinsic or extrinsic based abnormality can be identified. If the normal protein value of human is compare to lower than that of calculating cross ontology value (comparing BP&CC or MF&CC or MF&BP) is said to be Intrinsic. If the normal protein value of human is compare to higher than that of calculating cross ontology value (comparing BP&CC or MF&CC or MF&BP) is said to be extrinsic.

The relationship from various inputs are identified using MOAL (Multi ontology data mining at all levels) algorithm to mine the cross ontology. MOAL algorithm is to mine the cross-ontology based on association rules. The user can get the details about the gene id, extrinsic, intrinsic, disease, treatments on entering his / her gene report details.

3.4.MACHINE LEARNING ALGORITHM

Machine learning is an application of Artificial Intelligence (AI) that develops methods to automatically identify different patterns, apply hidden patterns to identify specific medicine and suggest tailored medicine to the patients. For training the model with conditions, we have used fuzzy based logic for training the models with promising accuracy. Thus personalized medicine is recommended for patients based on one's genomics lifestyle and other traits. Thus suggestion of tailored medicine for patients make better medicinal decisions reducing mortality rates.

4. EXPERIMENTS AND RESULTS

Fig 3 explains the end users the genes associated or related to each respective disease names. They can able to view complete ination like disease id, associated genes.

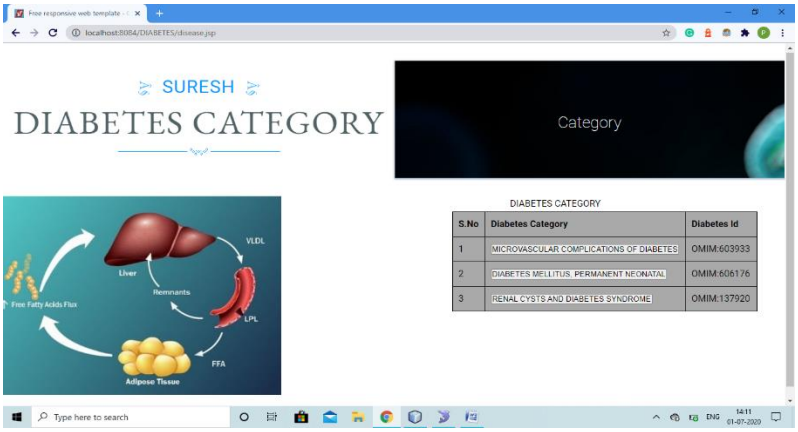

Fig 2. Disease names

Fig 2 explains the diabetics related disease names used in our proposed methodology for analysis.

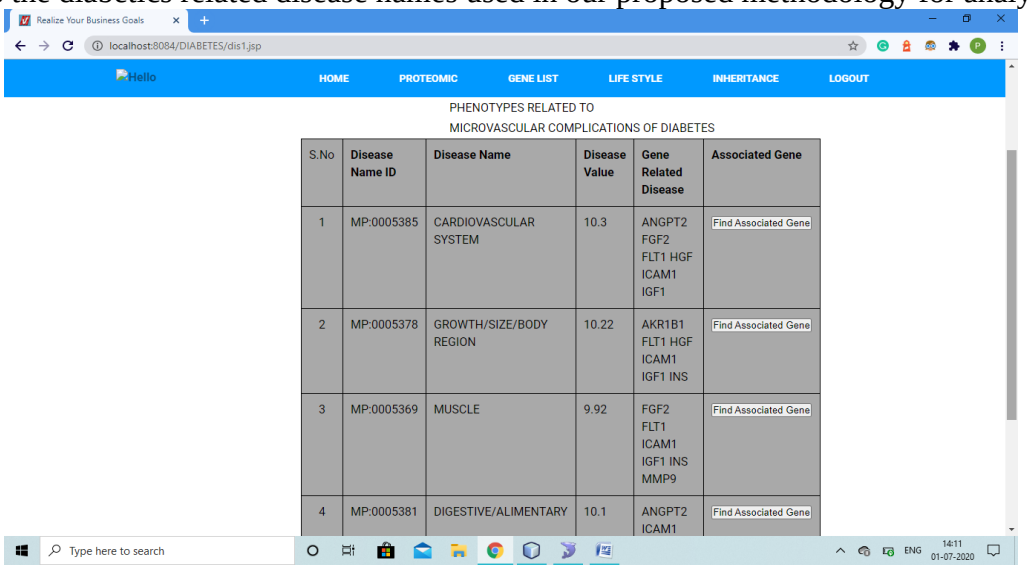

Fig 3. Gene related to disease

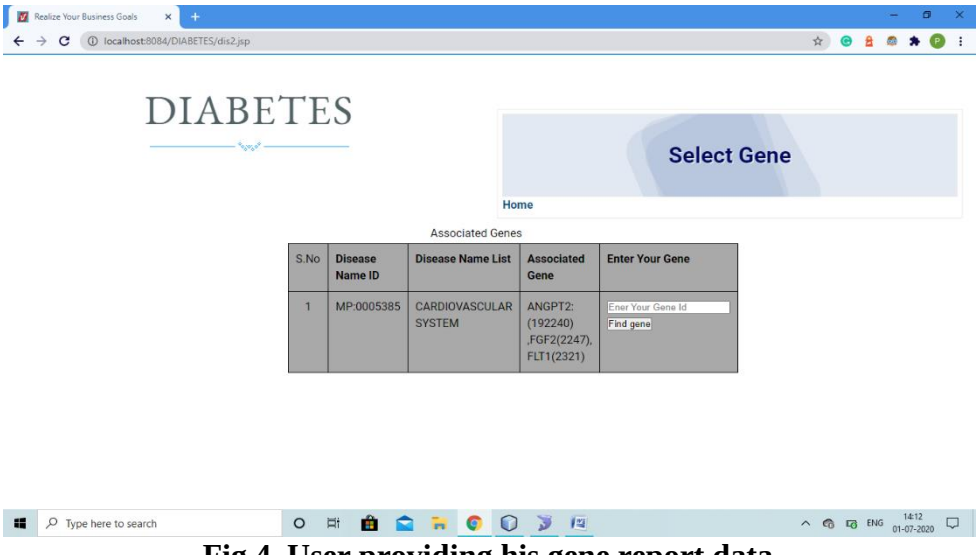

Fig 4. User providing his gene report data

Fig 4 is an important process, in which the users would provide their genetical test input values of enzymes under MF, BP, CC as stated in the test report. These values would be compared with the trained model for identifying which enzymes are extrinsic and intrinsic.

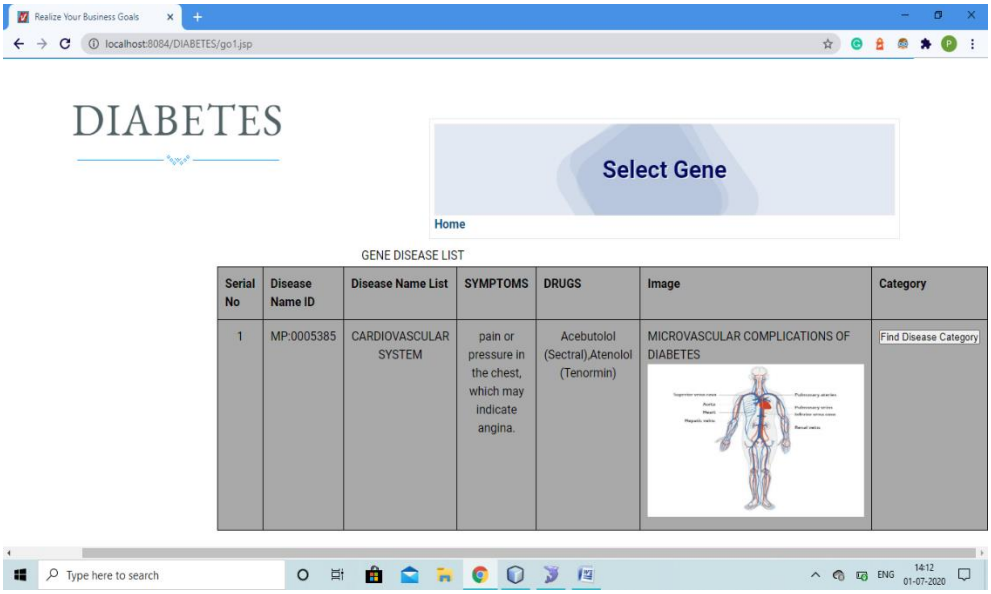

Fig 5. Based on the genetical input, intrinsic and extrinsic calculated and drug is suggested.

Fig 5 is the data analysis output in which the intrinsic and extrinsic enzymes are calculated on the scores entered by the user. Based on that medicinal drug is recommended.

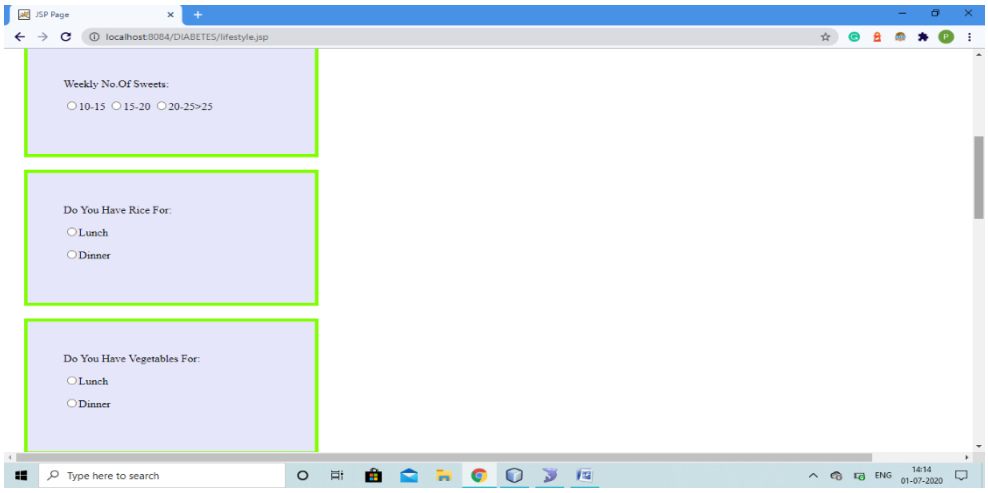

Fig 6. User lifestyle input

Fig 6 provides each user to answer few questions to understand their life style, inheritance, food habits, working nature of individuals before recommending the drug.

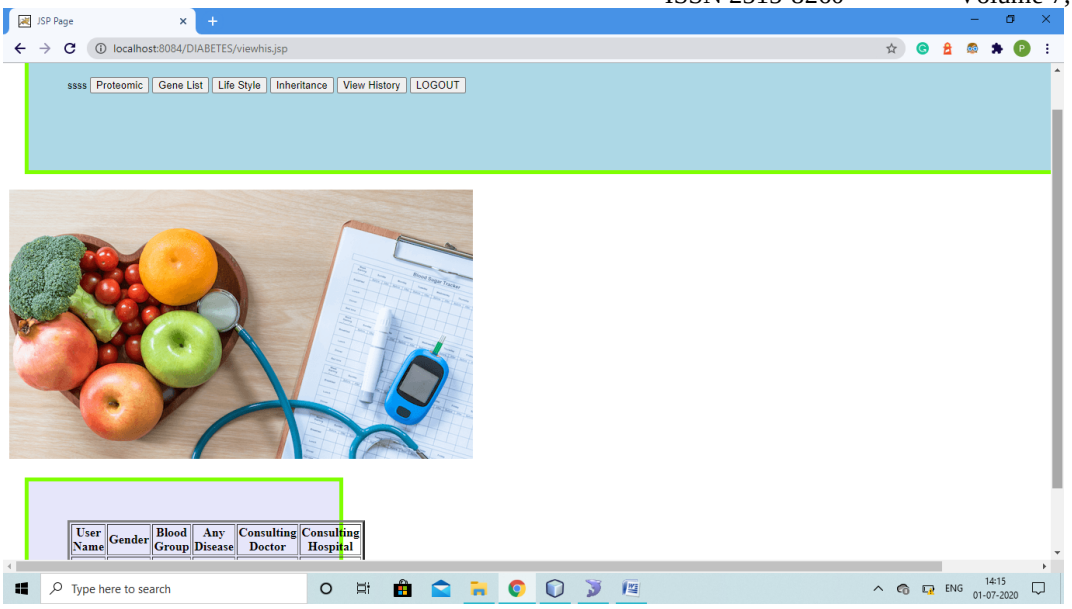

Fig 7. Personalized medicine is recommended

Fig 7 briefs based on the genetical, rna sequence, user lifestyle, food habits, working nature weightage value is defined by the proposed system based on the user inputs. Thus analyzing the weightage value, tailored medicine is recommended for each patients.

5. CONCLUSION

The development of personalized medicine would be an evolution in the healthcare sector. Thus this paper invokes personalized medicine recommendation based on several parameters such as age, weight, inheritance, genomic, proteomic, RNA sequences, life style, inheritance, food habitats provides an promising results. Thus this proposed system provides much advantages such as identification of disease at early stage, reduces mortality rate and improving prognosis.

6. REFERENCES

[1] A. J. Vargas and C. C. Harris, “Biomarker development in the precision medicine era: lung cancer as a case study,” Nature Reviews Cancer, vol. 16, no. 8, p. 525, 2016.

[2] Z. Wei, K. Wang, H.-Q. Qu, H. Zhang, J. Bradfield, C. Kim, E. Frackleton, C. Hou, J. T. Glessner, R. Chiavacci et al., “From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes,” PLoS genetics, vol. 5, no. 10, p. e1000678, 2009.

[3] J. Kim and H. Shin, “Breast cancer survivability prediction using labeled, unlabeled, and pseudo-labeled patient data,” Journal of the American Medical Informatics Association, vol. 20, no. 4, pp. 613–618, 2013.

[4] E.M.AntmanandJ.Loscalzo,“Precisionmedicineincardiology,” Nature Reviews Cardiology, vol. 13, no. 10, p. 591, 2016.

[5] D. Bzdok and A. Meyer-Lindenberg, “Machine learning for precision psychiatry: Opportunites and challenges,” Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2017.

[6] J. Davis, E. Lantz, D. Page, J. Struyf, P. Peissig, H. Vidaillet, and M. Caldwell, “Machine learning for personalized medicine: Will this drug give me a heart attack,” in the Proceedings of International Conference on Machine Learning (ICML), 2008.

[7] Kiruthika, U., Somasundaram, T.S. & Raja, S.K.S. Lifecycle Model of a Negotiation Agent: A Survey of Automated Negotiation Techniques. Group Decis Negot (2020). https://doi.org/10.1007/s10726-020-09704-z