

Biological Nutrient Deficiency Prediction using Machine Learning

Thirupathi Sundhararajulu¹, Yadamuru Rohini², Sourav Kumar³, K.Chaithanya kumar⁴,

A.Kunal Kumar⁵, A.M.Sai Krishnan⁶

^{1,2,3,4,5,6} *Computer Science and Information Technology, Siddharth Institute of Engineering & Technology*

Abstract- Deficiencies of nutrients including Iron, Vitamin C, Vitamin B12, and Vitamin D are common all over the world and they usually remain unnoticed until causing serious health complications. The conventional diagnostic procedures are costly clinical tests that restrict the ability to detect a condition early enough, particularly in under-served regions. This gap is filled in this project through creating a machine learning-based system that can predict nutrient deficiencies using common symptoms. The algorithms used in the system to give accurate predictions include Decision Tree, Random Forest, and XGBoost. The web application, which is developed with Flask, has a convenient interface and allows a person to enter their symptoms and receive the answer concerning possible nutrient deficiencies. Through the system, users have been enabled to make wise choices by having an affordable and easy way of tracking their nutritional health without necessarily undertaking costly medical tests. This preventive instrument of early detection increases health awareness, and promotes preventive health, which can lead to improved health outcomes of a population. The system is simple and easy to use by incorporating advanced machine learning methods that allow a high level of accuracy in prediction. The proposed solution will provide scalability to individual and healthcare provider users with an opportunity to intervene early and make an informed decision on nutritional health.

Keywords: Keywords Machine learning, nutrition, decision tree, XGBoost, random forest, nutrient deficiency, machine learning, tool (flask), machine learning, diagnosis of symptoms, machine learning, healthcare, early detection, nutrition, health care technology.

1. INTRODUCTION

Over the past years, the problem of nutrient deficiencies has become a major issue to the health of the world. Such deficiencies may be identified early to achieve

improved health outcomes by taking intervention. Nevertheless, these deficiencies may need clinical tests to be diagnosed and these tests may not be always available and convenient. The project will fill this gap by offering an automated platform which can predict deficiency of nutrients based on common symptoms. With the help of machine learning, the system can help those with possible nutrient problems to realize that they face a serious issue even before it becomes life-threatening. It is a strategy that can introduce cost-effective and real-time.

2. SYSTEM ANALYSIS

Existing System: Currently, most systems for diagnosing nutrient deficiencies involve laboratory tests or doctor consultations, which are often time-consuming and expensive. Some applications may provide symptom checkers, but they often lack the precision and accuracy required for reliable predictions. Many existing systems do not leverage advanced machine learning techniques for predictive analysis. Additionally, current tools do not offer personalized recommendations or real-time assessment. The majority of tools are limited to general health awareness and do not provide a specific diagnosis based on user input. There is a clear gap in providing an affordable and accessible system for predicting nutrient deficiencies.

Limitations of Existing Systems:

- High cost and inaccessibility of clinical testing.
- Lack of real-time and personalized prediction.
- Limited or no use of advanced machine learning for accuracy.
- Time-consuming processes that require visits to healthcare professionals.
- General health awareness tools with low precision and reliability.

3. PROPOSED SYSTEM

The project methodology will aim at creating a machine learning-based system that anticipates the nutrient deficiencies by observing the symptoms. This is done by collecting data, pre-processing it, training the model and finally developing an easy to access web application. The

system empowers machine learning algorithms like Decision Tree, Random Forest, and XGBoost to forecast the shortage of Iron, Vitamin C, Vitamin B12, and Vitamin D.

3.1 DATA COLLECTION AND PRE PROCESSING

The methodology involves first collecting an extensive amount of data on symptoms related to different deficiencies of nutrients. It is preferable that the dataset contains a comprehensive information like age, gender, lifestyle, and certain symptoms such as fatigue, weakness and dizziness that are typically associated with Iron, Vitamin C, Vitamin B12 and Vitamin D deficiency. In this project, a simulation of real-world data is done with a synthetic dataset or a publicly available dataset in the area of health and nutrition.

3.2 Model Selection and Training

The second thing is to train the machine learning models. This task is performed with the help of three major algorithms, Decision Tree, Random Forest, and XGBoost. Both algorithms are selected because they can be used to conduct classification and come up with interpretable outputs. The nutrient deficiency type (i.e. Iron, Vitamin C, Vitamin B12, or Vitamin D) is the target variable, and the dataset is preprocessed before being used to train the models.

1. Decision Tree: It is a simple algorithm but effective which is capable of creating a tree-like model of the decisions and their potential outcomes. It can be used to know the significance of various characteristics in predicting the nutrient deficiencies.
2. Random Forest: This is an ensemble learning algorithm which uses a combination of many decision trees to enhance accuracy and minimize overfitting. This model works well especially when dealing with a large number of features and interactions.
3. XGBoost: This is a gradient boosting algorithm that is characterized by a very high performance rate and capacity to work with complicated data. It is also very effective in enhancing the precision of the model since it reduces errors that may have occurred during previous iteration.

Advantages of Proposed System:

- Cost-effective compared to traditional clinical testing.
- Instant predictions based on symptom input.
- Machine learning-driven accuracy for better diagnosis.

- Accessible via a web application, no need for doctor visits.

EXPERIMENTAL RESULTS:

Diagram 1: System Architecture

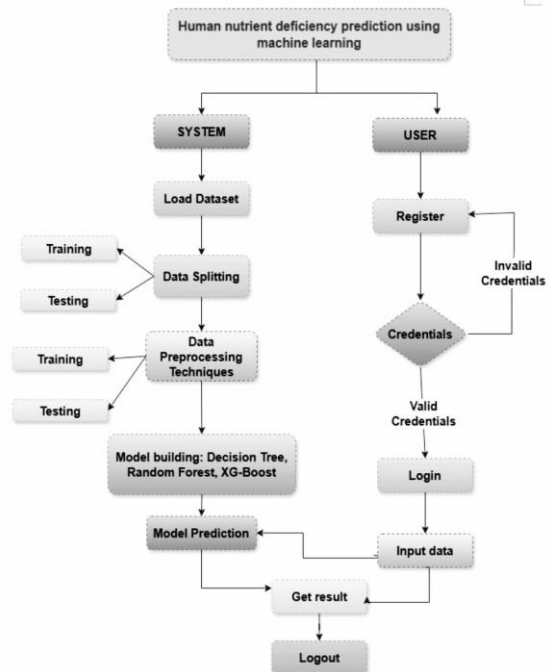


Fig -1: Figure

4. CONCLUSION

This project indicates how machine learning can be used to predict nutrient deficiencies, i.e. Iron, Vitamin C, Vitamin B12 and Vitamin D based on the common symptoms. The conventional ways of diagnosis can be expensive when it comes to clinical tests and thus early detection is not easily realized, particularly in the underserved regions. With the use of the Decision Tree, Random Forest, and XGBoost algorithms, the system is the efficient, cost-effective, and accessible way of detecting the nutrient deficiencies without having to conduct costly tests. The web-based application is created using Flask and enables users to enter the symptoms and receive fast feedback regarding the possible deficiencies of the nutrients. According to the outcome of the models, accuracy and precision are promising and therefore give reliable predictions.

FUTURE SCOPE

This project could be extended in future to contain more nutrient deficiencies like Vitamin A, Vitamin E and other necessary minerals. Also, the application of the methods of real-time data collection (i.e. mobile health app or wearable) would be beneficial to the user experience since they will get individualized recommendations due to ongoing tracking of symptoms. Moreover, partnerships with health professionals might result in

creating a more integrated health assessment tool, which will provide users with information on food-related changes and healthcare appointments.

REFERENCES

- [1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Mar. 2016, doi: 10.1145/2939672.2939785.
- [2] M. Fiaz, A. Mahmood, K. Y. Baek, S. S. Farooq, and S. K. Jung, "Improving Object Tracking by Added Noise and Channel Attention," *Sensors 2020, Vol. 20, Page 3780*, vol. 20, no. 13, p. 3780, Jul. 2020, doi: 10.3390/s20133780.
- [3] M. R. Kim and S. J. Jeong, "Relationship between Vitamin D Level and Lipid Profile in Non-Obese Children," *Metabolites*, vol. 9, no. 7, p. 125, Jul. 2019, doi: 10.3390/metabo9070125.
- [5] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Towards Medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Aug. 2020, doi: 10.1109/TNNLS.2020.3027314.
- [6] S. Willner, R. Whitemore, and D. Keene, "'Life or death': Experiences of insulin insecurity among adults with type 1 diabetes in the United States," *SSM Popul. Health*, vol. 11, p. 100624, Aug. 2020, doi: 10.1016/j.ssmph.2020.100624.

BIOGRAPHY



I, THIRUPATHI

SUNDARARAJULU, currently working as Assistant Professor in the Department of Computer Science & Information Technology at Siddharth Institute of Engineering & Technology,

Puttur, Andhra Pradesh, India. I am currently pursuing my Ph.D. in Computer Science and Engineering at Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai. My research interests include Deep learning, Bigdata, and Cybersecurity. I published research papers in faculty development programs, workshops, and technical seminars. I guided several undergraduate projects and committed to academic excellence and research innovation.