

**AI-BASED SYMPTOM-DRIVEN PREDICTION OF VITAMIN B12
DEFICIENCY RISK IN WOMEN: A COMPREHENSIVE REVIEW****Sonali R. Patil¹, Dr. Rupali H. Patil²**¹ *Research Scholar, Department of Computer Science, SSVPS Lt. Karmveer Dr. P. R. Ghogare Science College, Dhule.*Email: Sonalivmarathe65@gmail.com² *Professor, Department of Computer Science, SSVPS Lt. Karmveer Dr. P. R. Ghogare Science College, Dhule.***Abstract**

Vitamin B12 deficiency is a serious public health concern, particularly for women, as it is often linked to anemia, problems during pregnancy, and long-term neurological damage. While its dependability, traditional diagnostic approaches frequently experience delays due to the non-specific presentation of early symptoms and the resource-intensive nature of clinical testing. This study looks at current advances in artificial intelligence (AI) and machine learning (ML) for the symptom-driven prediction of vitamin B12 deficiency and related nutritional concerns in women. The review identifies a wide range of modeling techniques, such as supervised machine learning (Random Forest, Support Vector Machines, Logistic Regression), deep learning (Artificial Neural Networks), and innovative hybrid frameworks like Knowledge-Guided Neural Networks (KGNNs). Random Forest and ensemble techniques are routinely the strongest models for forecasting nutritional status and anemia severity, often exceeding 90% accuracy. Dietary consumption profiles, self-reported symptoms, and Complete Blood Count (CBC) parameters have been found as significant input features in the literature. An essential part of modern predictive systems is the incorporation of individualized dietary recommendation engines. Prototypes shown in recent studies indicate the practical application of turning AI-derived risk assessments into helpful food suggestions. However, the paper highlights an ongoing issue with clinical use: the lack of transparency of "black-box" models. It is shown that Explainable AI (XAI) techniques, such as SHAP and LIME, can bridge this gap by providing simple forecast explanations. Given these improvements, the study concludes that there is still an important lack of vitamin B12-specific frameworks created especially for women, which is an important topic for further study.

Keywords: Vitamin B12, AI(XAI), ML, Symptom, Deep Learning.► *Corresponding Author: Sonali R. Patil***1. Introduction**

Cobalamin, frequently referred as vitamin B12, is necessary for the production of DNA, red blood cells, and a healthy mental system. Lack of it is a major contributor to megaloblastic anemia and, if ignored, can result in irreversible nerve damage. Due to the physical demands of pregnancy and lactation, as well as the fact that some groups of people are more inclined following to strict diets (such as vegetarianism or veganism), women are especially susceptible to not gaining enough B12. [16], [17].

It might be difficult to identify particular symptoms early on as they lack in specificity and can be mistaken for other conditions including low iron levels and anemia. These symptoms include fatigue, weakness, and mental confusion. Blood B12 tests, which are necessary for traditional diagnosis, may not always be available in certain areas with limited resources. By analyzing dietary habits, lifestyle habits, and symptom patterns, artificial intelligence (AI) provides a novel approach to anticipate deficiency risk or use as a screening tool prior to clinical testing. [1], [15]. The current state of AI-based predictive algorithms for vitamin B12 and other vitamin and mineral deficiencies is examined in this paper, with a focus on how these models might be used to improve the health of women. It investigates how explainability affects model trust, how models have evolved from being only based on clinical data to being dependent on symptoms and lifestyle qualities, and how these systems can be utilized for giving individualized health advice.

2. Predictive Modeling Techniques

The literature describes a transition from traditional statistical methods to sophisticated machine learning architectures. These models are generally categorized into supervised learning, deep learning, and hybrid systems.

A change from simple statistical techniques to sophisticated machine learning frameworks is described in the literature. These models are typically divided into three categories: hybrid systems, deep learning, and supervised learning.

2.1 Supervised Machine Learning

For predicting dietary deficits, supervised learning remains the most important method. K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machines (SVM), and Logistic Regression (LR) are a few popular methods. Because Random Forest (RF) can handle non-linear relationships and multidimensional data, it is frequently claimed to be superior to other approaches. RF has been successfully used to predict nutritional deficits among women in Bangladesh and Pakistan [9], [17], and for feature selection in female sub-health situations [3].

Support Vector Machines (SVM): As SVM functions well in multi-dimensional situations, it is commonly used for problems with classification. SVM-based models with classifier chains carried out better on multi-label classification tasks, such as recognizing vitamin B12 and multiple kinds of anemia [15]. The group and Boosting Techniques Nowadays, advanced methods such as XGBoost and CatBoost are used to increase accuracy and speed implementation. XGBoost has been recognized as a top technique for anemia prediction in datasets that contain several elements related to iron and vitamin absence [10], [11].

2.2 Deep Learning and Artificial Neural Networks

Artificial Neural Networks (ANNs) can be utilized to match specific deficiency labels with complicated collections of symptoms. Neural networks were used by early expert systems with an accuracy rate of more than 60% to detect deficiencies and suggest diets based on general symptoms [1]. Large-scale diets are an example of more recent uses of deep learning, but they often find it difficult to be clear about what they do.

2.3 Hybrid and Knowledge-Guided Frameworks

An important development in the field is the growth of Knowledge-Guided Neural Networks (KGNNs). The neural network architecture of such models directly includes domain knowledge, like biochemical methods and nutritional ontologies [2]. In contrast with traditional "black-box" models, this method is more clear and has shown a high accuracy rate (93.7%) for detecting vitamin deficiencies. For people at-risk, KGNNs have become especially important as medical standards must be strictly observed to [2].

3. Data Preprocessing and Feature Selection fatigue, weakness, and mental confusion

Any AI model's performance depends upon the quality of the input data and the value of the selected features.

3.1 Symptom-Driven and Non-Clinical Data

The change to non-surgical testing resulted in a growth in the use of lifestyle and symptomatic data.

Symptoms: fatigue, weakness, and mental confusion and physical indications (such those in the hand or eye) are the primary markers [1], [16]. **Dietary Intake:** Data on the consumption of specific food groups or vitamin-rich products are used to forecast future deficiency risks [13], [19].

Lifestyle and Demographics: Research on women in underdeveloped countries has shown that age, geography, and socioeconomic level are significant risk markers [9], [17].

3.2 Clinical and Biochemical Parameters

Even though the goal is often symptom-driven prediction, clinical metrics, such as Complete Blood Count (CBC) data, are frequently used to validate models or give a hybrid input. Hemoglobin (Hb), mean corpuscular volume (MCV), and red blood cell (RBC) count are essential indicators to differentiate between a lack of iron (often microcytic) and vitamin B12 deficiency (generally macrocytic) [7], [11], and [15].

4. Findings and Comparative Analysis

4.1 Model Performance and Reliability

A comparative research regularly shows the performance of collaborative methods. For instance, in a study of Pakistani women, Random Forest outperformed SVM, Logistic Regression, and KNN [9]. Similar findings found that in Bangladesh, an RF-RF combination (feature selection and classifier) obtained 81.29% accuracy [17]. With CBC data, Random Forest and CatBoost achieved up to 99.2% accuracy in clinical situations [11].

Non-invasive methods are also promising. AI may utilize physical signs that link with fundamental vitamin status, as proved by a study that used KNN to evaluate photos of the palm and fingernails to detect anemia with an accuracy rate of 99.88% [16].

4.2 Hyperparameter Tuning and Optimization

By using techniques like Grid Search and Random Search for improving model parameters is an essential step in achieving high precision. Researchers have shown that enhancing the hyperparameters of ensemble models can boost accuracy to 96% or higher, verifying that the models are sufficient in reliability for clinical use [5], [12].

5. Integration with Personalized Dietary Recommendations

5.1 Intelligent Food Guidance Systems

The main objective of many predictive systems is to transform from diagnosis to control and preventive. Models that link dietary recommendations with deficiencies prediction have been proposed by many researchers.

Expert System Integration: From the beginning neural network models were created to offer nutritional and medicine suggestions following a diagnosis [1]. **Vitamin-Intake Based Recommenders:** Specific guidance systems have been designed to turn vitamin-status data into particular dietary dietary recommendations in order to assist users keep healthy nutrient levels [4], [19].

Mobile and Web Applications: The implementation using frameworks like Flask resulted in user-facing apps that predict gaps and suggest meal combinations [19].

5.2 Rule-Based vs. AI-Driven Suggestions

As earlier systems relied on simple rule-based mappings, recently developed hybrid models (KGNNs) use counterfactual analysis to develop recommendations. This results in it achievable for the system to propose specific, customized dietary changes that would most successfully solve a predicted gap [2].

6. Explainable AI (XAI) and Clinical Transparency

The “black-box” characteristics of multiple high-performing systems is a major barrier to the use of AI in healthcare. Transparency is being achieved through the integration of Explainable AI (XAI) approaches. SHAP and LIME: The post-hoc methods of interpretation are used to find out which variables (such as a small amount of red meat or specific symptoms) had a major effect on a given prediction [5], [6].

Attention and Concept Activation: Personal attention systems are used by hybrid models such as KGNNs to explain the importance of certain dietary mechanisms, providing a more “white-box” technique that doctors might count on [2], [6].

Transparency for Trust: By explaining why behind a diagnosis or a proposed dietary adjustment, transparent models have been proved to increase doctor acceptance and compliance among patients [6].

7. Research Gaps and Challenges

Despite the progress, several key challenges remain: Women Lack B12-Specific Frameworks: Anemia and multivitamin deficiencies are identified in recent studies. However, few machine learning systems are specifically created to analyze the different physiologic and nutritional risk characteristics of women who are lacking in vitamin B12 [b12_ai_insights. md].

Clinical Integration: Model development is still the only way to fully implement such models into real-world clinical procedures or intelligent food guidance systems, with the fact that we achieve good accuracy in research settings [b12_ai_insights. md].

Data Standardization: The growth of a standard cross-population predictor becomes challenging by the different use of lifestyle and clinical characteristics across research.

8. Conclusion

AI-based predictive modeling is an effective method for women to find and take care of vitamin B12 deficiency before on. By combining with customized nutritional advice systems and applying symptom-driven data, these devices can offer low-cost, non-invasive monitoring services. Although hybrid systems like KGNNs and group models like Random Forest are extremely accurate, explainable AI is required for coupling clinical utility and prediction accuracy. To improve long-term health results, future research should concentrate on developing B12-specific frameworks for women and formal intelligent food guidance systems.

References

1. N. Yadav, S. S. Sharan, and S. S. Prasad, "Vitamins and Minerals Diagnosis System Using Neural Network," in Proc. International Conference Inventive Communication and Computational Technologies (ICICCT), 2018. DOI: [10.1109/ICICCT.2018.8473019](https://doi.org/10.1109/ICICCT.2018.8473019)

2. T. Revesai et al., "Knowledge-Guided Neural Networks for Transparent Micronutrient Deficiency Detection," in Proc. ICTAS, 2025. DOI: [10.1109/ictas64866.2025.11155502](https://doi.org/10.1109/ictas64866.2025.11155502)
3. J. Wang et al., "Application of random forest data mining method to the feature selection for female sub-health state," in Proc. IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW), 2010. DOI: [10.1109/BIBMW.2010.5703880](https://doi.org/10.1109/BIBMW.2010.5703880)
4. S. Reddy, "Diet Recommendation System Based On Vitamin Intake," Advances in Computational Sciences and Technology, vol. 16, no. 1, pp. 35–43, 2023. DOI: [10.37622/acst/16.1.2023.35-43](https://doi.org/10.37622/acst/16.1.2023.35-43)
5. S. Kumar et al., "Anemia Detection And Severity Prediction Using Classification Algorithms With Optimized Hyperparameters, Boosting Techniques And XAI," in Proc. INCET, 2024. DOI: [10.1109/incet61516.2024.10593447](https://doi.org/10.1109/incet61516.2024.10593447)
6. M. S. Farooq et al., "Developing a Transparent Anaemia Prediction Model Empowered with Explainable Artificial Intelligence," IEEE Access, 2024. DOI: [10.1109/access.2024.3522080](https://doi.org/10.1109/access.2024.3522080)
7. "Application of Machine Learning Algorithm in Identification of Anaemia Diseases," in Advances in Computational Intelligence, Springer, 2022. DOI: [10.1007/978-981-19-3391-2_8](https://doi.org/10.1007/978-981-19-3391-2_8)
8. S. Routman, "Prediction of Anemia Disease Using Machine Learning Algorithms," in Proc. ICCS, 2023. DOI: [10.1007/978-981-99-0071-8_18](https://doi.org/10.1007/978-981-99-0071-8_18)
9. "Machine Learning Approaches for Prediction of Nutrition Deficiency among Women of Different Age Groups," in Proc. ICONICS, 2022. DOI: [10.1109/iconics56716.2022.10100340](https://doi.org/10.1109/iconics56716.2022.10100340)
10. Schweta et al., "Prediction of Anemia using various Ensemble Learning and Boosting Techniques," EAI Endorsed Transactions on Pervasive Health and Technology, vol. 9, 2023. DOI: [10.4108/eetpht.9.4197](https://doi.org/10.4108/eetpht.9.4197)
11. N. Aditya et al., "Machine Learning Models for Classification of Anemia from CBC Results: Random Forest, SVM, and Logistic Regression," INTI JOURNAL, 2024. DOI: [10.61453/jods.v2023no49](https://doi.org/10.61453/jods.v2023no49)
12. S. Rana et al., "Predicting Anemia with Precision: Employing Machine Learning Ensemble Modeling and Hyperparameter Tuning," in Proc. ICONAT, 2024. DOI: [10.1109/iconat61936.2024.10774701](https://doi.org/10.1109/iconat61936.2024.10774701)
13. S. Kosaraju et al., "A Model for Analysis of Diseases Based on Nutrition Deficiency Using Random Forest," in Proc. ICCES, 2022. DOI: [10.1109/icces54183.2022.9835805](https://doi.org/10.1109/icces54183.2022.9835805)
14. M. Ali et al., "Machine Learning Approaches for Prediction of Nutrition Deficiency among Women of Different Age Groups," Iconics, 2022. DOI: [10.1109/ICONICS56716.2022.10100340](https://doi.org/10.1109/ICONICS56716.2022.10100340)
15. J. Patel et al., "Anemia Multi-label Classification Based on Problem Transformation Methods," in Soft Computing and Signal Processing, Springer, 2021. DOI: [10.1007/978-981-16-0733-2_44](https://doi.org/10.1007/978-981-16-0733-2_44)

16. S. Aboul-Yazeed et al., "Non-Invasive Anemia Detection using Machine Learning: A Comparative Analysis," *Advanced Sciences and Technology Journal*, 2025. DOI: [10.21608/astj.2025.390972.1064](https://doi.org/10.21608/astj.2025.390972.1064)
17. S. Islam et al., "Risk factors identification and prediction of anemia among women in Bangladesh using machine learning techniques," *Current Women's Health Reviews*, vol. 17, 2021. DOI: [10.2174/1573404817666210215161108](https://doi.org/10.2174/1573404817666210215161108)
18. S. Sailaja, "Vitamin Deficiency and Food Recommendation Using Machine Learning," *International Journal for Research in Applied Science and Engineering Technology*, 2024. DOI: [10.22214/ijraset.2024.59279](https://doi.org/10.22214/ijraset.2024.59279)