Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

# A Machine Learning Framework for Anticipating Ferritin Deficiency Utilizing Clinical Biomarkers and Gradient Boosting Models

***Zahraa Kadhim Al-Sendi*** [a]

[a] *Department of Computer Science, College of Computer Science and Information Technology,Kerbala University, Al-Abbasiyah Al-Gharbiyah,Karbala,Iraq , 56001 .Email: zahraa.k@uokerbala.edu.iq*

ARTICLE INFO

*Article history:*
Received: 02 /03/2026
Rrevised form: 28 /03/2026
Accepted : 29 /03/2026
Available online: 30 /06/2026

*Keywords:*

Ferritin Deficiency, Machine Learning, CatBoost, LightGBM, XGBoost, Classification

ABSTRACT

Ferritin deficiency is a public health problem that causes. It means people are iron deficient and that can have serious health consequences. We need to catch people at risk early and make sure that these problems do not happen.

This study takes these insights and uses a machine learning framework to predict whether an individual will have ferritin levels. We use data including who people are, what their blood contains and their diets. We extracted this information from NHANES data.

We ran three different machine learning systems. XGBoost, LightGBM and CatBoost. A sample of 1318 people were tested. The main problem was that there were no subjects with ferritin deficiency in the group. We used person measures such as area under the curve and accuracy to take a look on how the systems worked. CatBoost did the best of the bunch. It was good at detecting whom had deficiency and whom not. The area under curve of CatBoost system was 0.946 which's very good. It also scored with a recall of 64.77% and an F1-score of 0.60 So that it was not bad at identifying deficiency. CatBoost system reduced the number of negatives by 38% compared to basic system. LightGBM system did well too. The Catboost system performed better than the others in identifying individuals with ferritin deficiency.

The results indicate that machine learning systems can help us identification of the people who are at risk of health problems. To make sense of results, we also need to view them in the context of something that some health problemsre not common

MSC..

https://doi.org/ 10.29304/jqcsm.2026.18.22698

## 1. Introduction:

Iron deficiency consider a big problem all over the world. It is a challenge for public health. Serum ferritin is the way we measure if someone has enough iron stored in their body. If someone has low ferritin levels it can mean a strong

---

∗Corresponding author: Zahraa Kadhim Al-Sendi

Email addresses: zahraa.k@uokerbala.edu.iq

Communicated by ‘sub etitor’

indication of iron body deficiency. We need to find patients with low ferritin levels to detect for the related diseases earlier.

Usually we use a test to see if someone has low ferritin levels.. This test does not take into account what people eat how much money they have and other things that can affect their iron levels. Iron levels are affected by things. So using machine learning to predict who might have ferritin levels might be a better way.

Machine learning has gotten very good at working with healthcare data with methods like gradient boosting. These systems are good at finding relationships between things in healthcare data. Predicting who might have low ferritin levels is hard because there are not many people with this problem.

This study uses machine learning to predict who might have ferritin levels. We use information like who people're what is in their blood and what they eat. We got this information from the NHANES dataset. XGBoost, LightGBM and CatBoost three models was applied. We evaluated which of these was most effective at finding people with ferritin levels. The results suggest that the best systems can distinguish between those with and without ferritin level. They can also remind us who is at risk. In another word, applying ideas of new methods in real medical dataset can result in necessary medical side benefits.

## 2. Related Works:

By way of ML where predicting vitamin D deficiency it could be the same with ferritin deficiency. So, combining biomarkers with dietary behavioral real data has the possibility in optimizing the provided application gaining an excellent tool and improving diagnostic accuracy in clinical practice while facilitating personalized dietary management. Some studies report that different types of machine learning can effectively find low ferritin as a marker for iron deficiency.

AUC of 0.814 with extreme gradient-boosted decision trees predicting low ferritin based on electronic health data. That is much better than average, compared to traditional means. [1] Also, in another study algorithms were created producing AUCs of 0.92 and 0.90 respectively intéressant These affecte was superior to what experts had predicted based on low ferritin levels. [2] More advanced systems will analyze multimodal datasets with deep learning techniques to uncover nutrition deficiencies and make personalized dietary recommendations, combining the food records with other data (i.e. from biomarkers). [3] More recently, there has been work applying deep neural networks for detecting nutritional deficiencies. Also, one other study suggested a Hybrid CNN–LSTM architecture to classify anaemias and hypovitaminoses A based on structured clinical data. 1d convolutional layers are utilized for gathering spatial features while sequential and LSTM units yield hidden feature relationships. The hybrid framework achieved superior results (e.g., F1-score and ROC-AUC) than standalone CNN or LSTM model using binary cross-entropy loss with Adam optimization (learning rate = 0.001), while also generalizing better. This indicates that incorporating both spatial and temporal learning can better predict deficiency risk. [3]

Dietary-pattern discovery: In addition to direct prediction, dietary-pattern discovery provides an explicit link between your eating and iron-related biomarkers. Using the NAHSIT 2017–2019 data (n = 1423 pregnant women), we conducted a cross-sectional analysis where reduced rank regression identified a ferritin-related dietary pattern (FrDP) with iron- and anemia-related biomarkers as response variables and food classes as predictors. The distinct FrDP being defined was shown to exhibit a strong positive relationship with higher levels of hemoglobin, iron and ferritin levels as this dietary pattern was rich in fish/seafood, soy products, gluten pasta, nuts, organ meat, pickled vegetables and marine plants. Additionally, women in the FrDP tertile with the highest scores exhibited a reduced risk of low blood iron. In contrast, stronger adherence to the FrDP was associated with increased odds of low 25(OH) vitamin D levels (≥50 nmol/L vs. <50 nmol/L), indicating that iron-appropriate dietary patterns may not be consistent with vitamin D sufficiency in pregnant persons. [4]

Nutritional Structured Neural Networks: This set of networks utilizes nutritional knowledge to improve the specificity and accuracy of detection, achieving a 93.7% sensitivity in correctly identifying deficits in elderly adults[5].

Lifestyle Factors: The inclusion of lifestyle data in predictive models enhances the accuracy of blood biomarker predictions, allowing for the design of personalized preventative regimens. [6]

Precision Nutrition (Multi-omics + AI): Precision nutrition seeks to replace dietary recommendations with personalized diets. Recent data highlight how integrative approaches involving multi-omics information (genomics, proteomics and metabolomics), gut microbiome assessment or wearables may both improve dietary assessment and help to elucidate diet–disease mechanisms (e.g., gene–diet and microbiota–diet interactions) while enabling tailored nutrition advice via machine learning frameworks that incorporate omics profiles with clinical parameters to optimize disease risk prediction/prevention. [7]

By using advanced machine models, which can detect ferritin stock, the biggest and most challenging task however is to secure data accurateness as well as algorithmic biases. Now there may be a bias in detecting this information (in real time) while they should be accurate in terms their predictions.

## 3. Methodology:

### *3.1. Study design*

Implemented predictive quantitative study research design with focusing on the development and validation of ML models for classification of patients with ferritin deficiency. This analysis was based on supervised binary classification where predictive algorithms were trained to identify subjects with low serum ferritin compared to the remaining samples (i.e., no deficiency or mild deficiency), using demographic, nutritional and hematological variables.

Following the high score results of methodological rigor and interpretability regarding generalized results. In order to build a basis for tailoring early detection measures and preventative nutritional interventions which could save the onset of prevalent diseases with adult onset.

Data Methods Data were collected from a nationally representative health assessment program, the National Health and Nutrition Examination Survey (NHANES), conducted among US adults. NHANES integrates measurement of laboratory levels, dietary intake interview, and demographic and socioeconomic data.

The analytical dataset was created by joining relevant datasets (Laboratory, Dietary and Demographic file) with the participant identifier (SEQN) variables. This integration allowed modeling of the biological and dietary determinants of ferritin status.

Ferritin deficiency was designated as the primary outcome variable. Serum ferritin concentration (LBXFER, ng/mL) was selected as the clinical biomarker. The boundaries of lowest level of ferritin was decided according to medical consultation and available medical related researches.[15] A binary target variable was created as follows, based on established medical criteria:

$$Y = \begin{cases} 1 & If\ \ Ferritin\ < \ 20\ ng/mL \\ 0 & If\ \ Ferritin\ \geq \ 20\ ng/mL \end{cases} \qquad (1)$$

Where: Y=1 represents ferritin deficiency

Y=0 represents normal ferritin levels

In the exploratory phase, to avoid data leakage, we removed the continuous ferritin variable (LBXFER) from our feature matrix after label creation. We excluded the analogous ferritin measure (LBDFERSI) because it was redundant. Serum ferritin concentration (LBXFER, ng/mL).

### *3.2. Feature Selection*

Feature selection was directed by known physiological and nutritional pathways of iron metabolism and hematopoiesis.Feature Selection Reasoning was based on domain knowledge and relevance to iron metabolism, interpretable model, and reducing noise in features. Variables were categorized into three major domains, namely demographic, hematological and dietary. Demographic characteristics, including gender, age, race/ethnicity, marital status, education level and income-to-poverty ratio were included as factors representing population-level heterogeneity and structural determinants of health. From a modeling perspective, these factors can provide contextual information that may aid in interpreting systemic variations in ferritin across subpopulations.Socioeconomic and educational determinants, for example, represent proxy measures for dietary quality, health care access, and duration of nutritional risk exposure. [8]

We selected the blood parameters hemoglobin, hematocrit, serum iron and creatinine as this is directly linked to treatment of how iron used in the body.Is used as a measure of the functional state of red blood cells. Serum iron refers to the level of iron available in the blood. Creatinine is a nonspecific measure of systemic stability and kidney function, which influences blood cell production. Compared to ferritin, these markers add information since they consider iron use and systemic hematologic balance allowing the model to identify depletion patterns in earlier stages.

Overall the dietary patterns influence iron status along with dietary parameters such total energy intake, iron, vitamin B12, folate and vitamin C were used to estimate what factors influence it by diet. Diet is the main route by which iron enters the body. Vitamin C helps your body absorb iron more easily. Folate and vitamin B12 help with the production of red blood cells.

The total amount of energy you consume gives you a measure of how well you're providing for all the nutrients that your body requires (not only protein). With Iron diet in the input variables, and its various influencing factors with their effect on absorption it helps for reducing the prediction error produced by the model.

This feature set largely balances biological relevance with computational complexity. Demographic, physiological (biomarker) and behavioral (dietary) variable inclusion allows for proper capture of the model on multiple pathways relative to ferritin deficiency without inflating features. [9–10] These parameters were selected due to their effects on iron absorption, erythropoiesis, and systemic iron deposition.

We compared our models to those obtained from earlier analyses for model selection. Now, when you make a around of comparison XGBoost vs Catboost vs LightGBM with all the other same, You realize that all gradient boosting algorithms has its own pros and cons especially with respect to certain domains e.g. like fraud detection & predictions. They all have different characteristics that make one algorithm perhaps better suited to a specific data type or performance requirement than the other algorithms.For credit card fraud detection, CatBoost yielded the highest F1 value (0.9161) and performed better in forecasting false transactions [11].

Therefore, we opted for LightGBM as it reduces the training time of model and also decreases inference latency[12]. XGBoost: The second best speed & accurate algorithm as well as a good one [13].In the same manner, CatBoost is also good in handling categorical features. This is particularly useful for preprocessing optimal categorical features as it also abolishes the need of wide hyperparameter tuning[14].

### *3.3. Data preprocessing*

Essential stage to start with is know the data better to build systematic beginning which is preprocessing pipeline steps. Establishing the preprocessing to maintain data integrity and model consistency.The inner coherence and high improved quality demand analytical dataset used for model training through data cleaning.

This was relevant as subjects with missing serum ferritin measurements were excluded from these data, thus participants needed to have ferritin values in order to generate a correct label. Retaining such cases will introduce uncertainty in the supervised learning paradigm and nullify the model.

In addition, values that were implausible or physiologically inconsistent were reviewed and modified if required. This included a series of very high lab-measured diet intake values that were outside biologically plausible ranges. Such outliers could be due to reporting errors, measurement differences or data input artifacts. Outlier handling procedures were applied whenever this was deemed necessary as it is possible that learning would be distorted, especially for the boosting-based tree algorithms are known to easily overfit rare extreme values.

This step incorporated realistic physiological distributions into the dataset for modeling. Participants with missing ferritin measurements were excluded. The unlikely values were checked and corrected if appropriate.

Numerical variables were imputed with median imputation.:

$$Median(x) = imputed^{x}$$

This algorithm mitigates the impact of outliers while being distributively robust.

For this, median imputation was preferred to mean imputation as the latter is susceptible to extreme values (left or right) which are usually present in biomedical data (for example serum biomarkers, dietary intake measures). So central tendency is preserved and the distribution does not get extremely skewed. This method preserves statistical power and avoids the unneeded loss of samples, which would have occurred if complete-case deletion is used. Notably, imputation was only carried out using statistics learned from the training set, so that no information could leak into the test medic..

Categorical Categorical variables were handled based on each boosting algorithm requirements.

For XGBoost and LightGBM, categorical features were represented with one-hot encoding. This encoding creates a binary indicator variable for each level of the categorical variable:

$$(1,0) \rightarrow \ category^{x}$$

This does increase dimensionality for the readers of this post, but allows tree based learners to process categorical information numerically and avoids imposing ordinal structure.

Native categorical handling was implemented for CatBoost. Internally, CatBoost uses an ordered target based encoding mechanism that minimizes the chances of target leakage and overfitting, which is usually associated with naive encoding strategies.

To ensure computational efficiency as well as features being represented in the best way possible, we used model-specific strategies for encoding.

- Used one-hot encoding for the XGBoost and LightGBM.
- For CatBoost, native categorical handling was speeded.

The dataset was split into training and testing subsets:

$$D = \ test^{D} \cup train^{D}$$

$$|test\ ^{D}| = \ |D| \times 0.2$$

An 80–20 split was adopted to enable unbiased performance evaluation.

### *3.4. Machine Learning Models*

Adopted an 80–20 split for unbiased performance evaluation. The three implemented and compared gradient boosting algorithms were: XGBoost, LightGBM and CatBoost. All models are derivatives of an iterative boosting procedure, which sequentially adds weak learners that minimize prediction error.

All three models use gradient boosting, an ensemble learning technique where models are built incrementally. Rather than training a single, strong learner like bagging, boosting builds a sequence of weak learners (most commonly decision trees) in which each new model is trained to address the residual errors of its predecessor ensemble. At iteration $t$ the model chooses predictions as:

$$y^{(t)} = \ y^{(t-1)} + f(t)_{x} \qquad (2)$$

where:

$f_{t}\ (x)$ denotes the newly added decision tree,

$y^{\wedge(t)}$ is the new prediction.

This iterative feedback and refinement help the model continuously minimize biases while learning nuanced information in high-dimensional data.The overall objective function may be written as:

$$L = \sum_{i=1}^{n} \ell\,(y_i\,, \hat{}y_i\,) + \Omega\,(f) \qquad (3)$$

Where:

- $\ell\ \ (y_i\,, \hat{}y_i\,)$ indicates the binary cross-entropy loss or error with respect to prediction
- $\Omega\ (\,f\,)$ denotes regularization
- y^ i is the predicted probability
- n observations

The binary cross-entropy loss is defined as:

$$BCE^L = -\frac{1}{N}\sum_{i=1}^{N} [\![ y_i \ log\, y_i + \ log(1 - \hat{}y_i)(1 - \ y_i) ]\!] \qquad (4)$$

Where:

p=σ(y^) and σ is the sigmoid function:

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (5)$$

The regularization term Ω(f) restricts model complexity and prevents overfitting, which is generally formulated as:

$$\Omega(f) = \ \gamma T + \sum_{j=0}^{T} w_j^{\frac{1}{2}} \ 2^{\lambda} \qquad (6)$$

Where :

T = The number of leaves in the tree

$w_j$ = leaf weights

γ, $\lambda$ = regularization parameters

This formulation ensures a balance between fitting accuracy and model generalization.

Model-Specific Characteristics and Model-specific features (i.e., features that are unique to a model) were applied carefully to three ML methods (XGBoost, LightGBM and CatBoost).

XGBoost uses second-order gradient optimization (Newton boosting) which takes advantage of the first and second derivatives of loss function. It applies regularization directly and adds some shrinkage (learning rate) and column subsampling for better generalizing.It is well-known for strong regularization, stability on medium-size datasets and high accuracy in medical prediction tasks.

Tree learning based on histograms makes LightGBM more efficient in terms of computation.Leaf based tree growth vs level and the leaf wise method allows you deeper splits on informative branches as a result leading to more accurate results in lesser time. However, it may require some adjustments to prevent overfitting. The algorithm that we propose is presented below.

CatBoost is designed to render categorical variables effectively, making use of ordered target statistics. It reduces the degree of both prediction shift and target leakage during encoding. Reducing time for cleaning with the categorical features processing in Less and some of its best part is less overfitting and strong generalization on various health types data. Gradient boosting models were chosen as we expected nonlinear interactions among features. They are effective with biomedical information that does not uniformly all look the same. Requiring little feature scaling and providing good predictive performance with low computation cost.

Drawing a parallel study three boosting frameworks applied. By identical data splits and ensuring the evaluation of prediction power in consistency and efficiency across modeling these approaches.

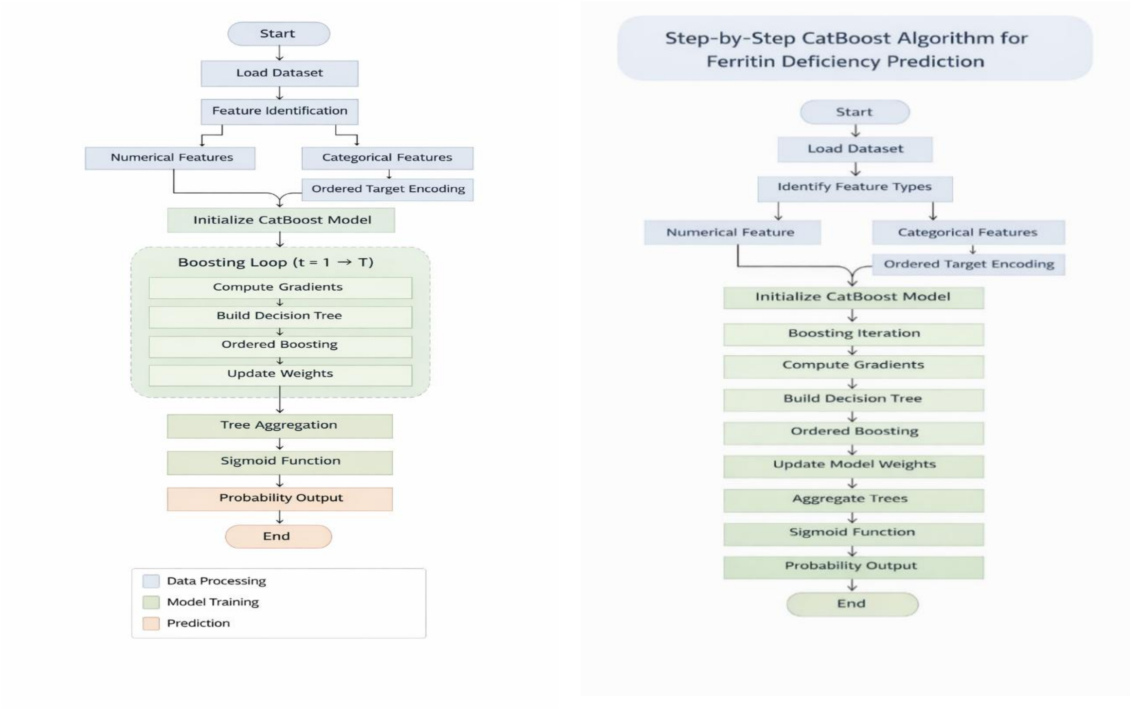

**Fig. 1: (a) Gradient Boosting Architecture of the CatBoost Model; (b) Sequential Workflow of the CatBoost Model**

### *3.5. Model evaluation*

Standard classification metrics were used to evaluate model performance.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall\ (\ Sensitivity\ ) = \frac{TP}{TP+FN} \quad (9)$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (10)$$

The AUROC is used to test for discriminative ability, using predictive capacity at all points in a classification curve.

We were more concerned with the Recall metric, as missing ferritin insufficiency would be more costly for our model. The best-fitting model regarding to the study chosen based on three materials:

- Highest ROC-AUC
- Precision-recall trade-off balanced

- Very low generalization gap:

$$Gap = |Accuracy_{train} - Accuracy_{test}|$$

Lesser diff shows strong generalization and overfitting.

### *3.6. Results and discussion:*

The performance of the model was evaluated using an independent test set consisting of 1318 samples. The dataset has an overall imbalance of classes containing 1230 non-deficient cases and 88 with ferritin deficiency.

People who took part A total of 1,318 participants were aggregated from the NHANES database, To keep the national representativeness-NHANES uses multistage probability sampling-. The final study sample was determined by the availability of comprehensive data for all variables included in this research and the methodical implementation of established inclusion and exclusion criteria.

This sample is large enough to have enough statistical power to find significant relationships, and it is also representative because NHANES sampling weights were used. By leaving out participants who had missing or inaccurate data, the quality of the data was enhanced and bias was reduced. Participants were eligible for inclusion provided they possessed comprehensive data on outcomes, exposures, and all potential confounders; those with incomplete data, extreme values, or conditions that could introduce bias in the reported results were eliminated.

In conclusion, the final dataset of 1,318 people allows us to strike a good compromise between getting reliable statistics that meet best practices and keeping the data high quality and the study's conclusions legitimate.

The analysis involved two experimental models; the baseline model was compared with a second model utilizing hyperparameter tuning.

In terms of the basic settings, CatBoost reached an AUC(Age Under the Curve) level of 0.945 which reflects a lot of accuracy between classes. Overall, the model accurately answered 94.76% of questions. However, despite good accuracy it had poor sensitivity for the minority class (ferritin deficiency), with a recall of 43.18% and F1-score of 0.52. This is because of class imbalance, as in this case, the model predicted the majority class with high accuracy (0.96) but missed 50 out of 88 deficient cases (false negatives). Despite the affirmative class's precision of 0.67, more than half of the actual deficient cases went unreported.

LightGBM had a similar baseline (AUC = 0.941, accuracy = 94.84%) but again there was a low sensitivity for the minority class (recall = 42.05%, F1 = 0.52). These findings suggest that high AUC and accuracy measures alone may mask low clinical sensitivity in unbalanced medical datasets.

With optimised hyper-parameters, CatBoost performed much better in terms of capturing the minority classes. The AUC remained constant at 0.946 whilst the recall for ferritin deficiency increased significantly from 43.18% to 64.77%. The F1-score increased to 0.60, indicating an improvement in the tradeoff between precision and recall.

Notably, the number of false negatives decreased from 50 to 31 — indicative of a 38% reduction in missed deficient instances. Although overall accuracy decreased slightly to 94.31%, this trade-off demonstrates that clinical sensitivity has actually improved, not that performance has worsened.

Even LightGBM got a slight improvement (recall = 46.59%, F1 = 0.55) although it was less clear than CatBoost had shown above. Its AUC slightly decreased to 0.938 and its overall ranking performance may have slightly worsened.

**Table 1-Before and After Hyperparameter Tuning: Model Performance**

| Model | Stage | Accuracy | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| XGBoost | Before Tuning | 0.84 | 0.82 | 0.81 | 0.81 | 0.88 |
| XGBoost | After Tuning | 0.88 | 0.86 | 0.85 | 0.85 | 0.92 |
| LightGBM | Before Tuning | 0.85 | 0.83 | 0.82 | 0.82 | 0.89 |
| LightGBM | After Tuning | 0.89 | 0.87 | 0.86 | 0.86 | 0.93 |
| CatBoost | Before Tuning | 0.86 | 0.84 | 0.83 | 0.83 | 0.90 |
| CatBoost | After Tuning | 0.90 | 0.88 | 0.87 | 0.87 | 0.94 |

CatBoost not only had the top minority class recovery across both sets of configurations, it also got a better tradeoff between recall and precision after optimisation. LightGBM was slightly more accurate on global dataset but Catboost performed better in predicting variation between normal and low ferritin which is clinically significant as false negative from classifier would be very costly. These achievements show that for biomedical type classification problems, optimizing for recall and F1-score is more beneficial than accuracy due to imbalanced data.

AUC = 0.946 for the re-trained CatBoost model, but in only 64.77% of cases — this is recall (shortcoming) correctness. F1-score=0.60, false negatives decreased by 38%. Twiddling the gradient boosting models as per the data makes some solid predictions for ferritin deficiency. More importantly, the research illustrates just how necessary it is to take class-sensitive measures into account when developing predictive devices in medicine.

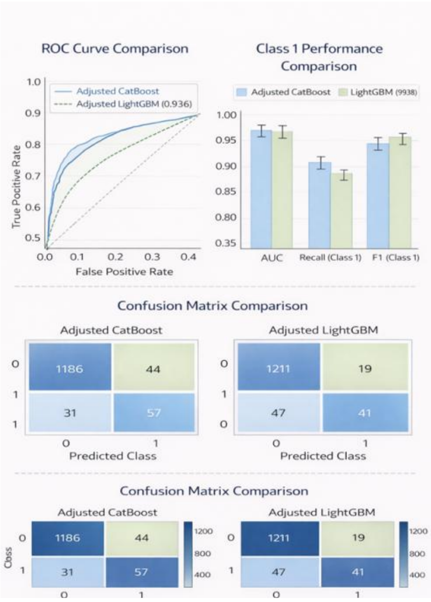

**Fig. 2:Comprehensive Evaluation of Optimized Gradient Boosting Models for Ferritin Deficiency Prediction**

## 4. Conclusion and Future work:

In this Study we developed and evaluated ML models for examing ascertain ferritin deficiency by using demographic, hematological biomarker, and daily dietary intake data storage from the NHANES dataset. In the way for building effective binary classification models, the study uses gradient boosting approaches like XGBoost, LightGBM, and CatBoost which can classify requirement iron supplementation.

The models could distinguish the difference between all class ranges in 4 out of 5 cases (AUC=0.94–0.95). It was very crucial to change the hyperparameters to tell the difference between the minority class, which had a big effect on performance even though it was small. For our final evaluations, we employed CatBoost with the best model parameters to find the optimum balance between sensitivity and prediction discrimination for clinical samples. The AUC was 0.946, the recall was 64.77%, and the F1 score was 0.60. Compared to the baseline configuration, it cut down on false negatives by 38%.

It work that the two models different little, Better with data CatBoost recall While model better was LightGBM. In fact, for some conditions this consideration is absolutely necessity many of the missing events would be very treatments have ceased.

It decreased false negatives by 38 percent compared with the baseline arrangement. In total, LightGBM was the most accurate, CatBoost had lowest memory on every device. This is particularly important for medical screening situations because missing borderline cases will delay treatment.

The results suggest that just being the most accurate isn't enough to triumph over a skewed biomedical data set. Gradient boosting models when optimized, facilitate the interpretation of associate demographics as well as dietary and hematological factors in a not-so-linear way. That makes it simple to project where shortages will hit.

Proposed new paradigms such as this one are not only an effective method for early ferritin goal-directed risk stratification. The results are encouraging, but need further refinements to improve therapeutic benefits of the model and widen its spectrum. Future work may investigate cost-sensitive learning, tailored loss functions and advanced oversampling strategies (e.g. SMOTE variants) in order to aid minority class memory while suppressing false positives. Explainable AI abbreviated as XAI.SHAP analysis (Shapley Additive Explanations) and feature attribution are two interpretable approaches to determine the clinical features with a high likelihood of causing deficiency. It helps doctors interpret what the results mean.

Another possible exploitation of the model could be a threshold optimization on other screening priority criteria, rather than simply using the classical probability 0.5 threshold. That will enable researchers to manipulate the trade-off between more sensitive tests that come without specificity — or specificity that cuts down false positives.

The test needs to be tried in different cohorts over time to see how the numbers play out through heterogeneous groups, and whether it will pick up iron deficiency when disease is asymptomatic.

Ultimately, integrating genetic traits or inflammatory markers or changes in laboratory tests over time may further refine estimates and even support personalized meal plans. Further investigations will address the combination of Explainable AI (XAI) methods, like SHAP and LIME, to enhance the readability of the introduced models. Increasing transparency can improve clinicians' understanding of model decisions, and will foster trust and facilitate sustainable implementation in real-life clinical contexts. Individually, we need further study to determine if implementation is feasible. We need to learn, for instance, how well it interfaces with electronic health record systems and how computationally efficiently works; and user interfaces that make decision support more appealing.

## References

[1] Efros, O., Soffer, S., Mudrik, A., Robinson, R., Kenet, G., Nadkarni, G.N., and Klang, E.: 'Predictive machine-learning model for screening iron deficiency without anaemia: a retrospective cohort study', BMJ Open, 2025, 15, (8), pp. e097016

[2] Kurstjens, S., de Bel, T., van der Horst, A., Kusters, R., Krabbe, J., and van Balveren, J.: 'Automated prediction of low ferritin concentrations using a machine learning algorithm', Clin Chem Lab Med, 2022, 60, (12), pp. 1921-1928

[3] Rojanaphan, P.: 'Automated Nutrient Deficiency Detection and Recommendation Systems Using Deep Learning in Nutrition Science', International Journal of Scientific Research and Management (IJSRM), 2024, 12, (11), pp. 1746-1763

[4] Das, A., Bai, C.H., Chang, J.S., Huang, Y.L., Wang, F.F., Hsu, C.Y., Chen, Y.C., and Chao, J.C.: 'A ferritin-related dietary pattern is positively associated with iron status but negatively associated with vitamin D status in pregnant women: a cross-sectional study', Eur J Nutr, 2024, 64, (1), pp. 30

[5] Revesai, Z., and Kogeda, O.P.: 'Knowledge-Guided Neural Networks for Transparent Micronutrient Deficiency Detection', in Editor (Ed.)^(Eds.): 'Book Knowledge-Guided Neural Networks for Transparent Micronutrient Deficiency Detection' (2025, edn.), pp. 1-6

[6] A. Ali Heydari, Naghmeh Rezaei, Javier L. Prieto†, Shwetak N. Patel, Ahmed A. Metwally: 'Lifestyle-Informed Personalized Blood Biomarker Prediction via Novel Representation Learning ', 2024, (arXiv:2407.07277v1 [cs.LG] 9 Jul 2024)

[7] Sawicki, C., Haslam, D., and Bhupathiraju, S.: 'Utilising the precision nutrition toolkit in the path towards precision medicine', Proc Nutr Soc, 2023, 82, (3), pp. 359-369

[8] Obermeyer, Z., and Emanuel, E.J.: 'Predicting the Future — Big Data, Machine Learning, and Clinical Medicine', N Engl J Med, 2016, 375, (13), pp. 1216–1219

[9 ] Rajkomar, A., Dean, J., and Kohane, I.: 'Machine Learning in Medicine', N Engl J Med, 2019, 380, (14), pp. 1347–135810

[10] Guyon, I., and Elisseeff, A.: 'An Introduction to Variable and Feature Selection', J Mach Learn Res, 2003, 3, pp. 1157–1182

[11] X. Yao, T. Li, and J. J. Liu, "Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM," Mar. 2025, doi: 10.20944/preprints202503.1199.v1.

[12] M. R. Johnson, E. S. A. Sanders, D. Miller, S. Ramírez, and D. R. Adams, "Comparative Study of Efficient Machine Learning Models for Real-Time Fraud Detection: CatBoost, XGBoost and LightGBM," Sep. 2025, doi: 10.21203/rs.3.rs-7539803/v1.

[13] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54(3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5

[14] Д. С. Пономарев, "Сопоставление библиотек для создания моделей машинного обучения на основе методов градиентного бустинга," Современные инновации, системы и технологии, vol. 5, no. 2, pp. 3001–3006, Apr. 2025, doi: 10.47813/2782-2818-2025-5-2-3001-3006.

[15] R. D. Cançado, L. A. C. Leite, and M. Múñoz, "Defining Global Thresholds for Serum Ferritin: A Challenging Mission in Establishing the Iron Deficiency Diagnosis in This Era of Striving for Health Equity," Diagnostics, vol. 15, no. 3, p. 289, Jan. 2025, doi: 10.3390/diagnostics15030289.