International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 318-329
https://doi.org/10.70917/ijcisim-2026-1761
Article

Leveraging Machine Learning and Calcium Metabolism Indicators for Predicting Calcium Homeostasis Disorders

Mohamed Attia[1], Ahmed M. Abd-Elwahab [2], Yehia Helmy [3], Hesham Mahmoud Ibrahim [4] and Mahmoud M. Bahloul [5]

[1] Future University in Egypt, Cairo, Egypt
[2,3,5] Business Information Systems, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt
[4] Business Information Systems Department, Modern Academy for Computer Science and Management Technology, Cairo, Egypt.
[1] Mohamed.abdelgwad@fue.edu.eg, [2] ahmed.mohamed.abdelwahab@commerce.helwan.edu.eg,
[3] ymhelmy@yahoo.com, [4] dr.heshammahmoud122@gmail.com and [5] mahmoud.bahloul@commerce.helwan.edu.eg
Corresponding Author: Mohamed.abdelgwad@fue.edu.eg

Abstract: Many physiological systems depend on calcium homeostasis, which can be disrupted to cause conditions like hypercalcemia, hypocalcemia, and hyperparathyroidism. Effective management and treatment of many illnesses depend on an early and precise diagnosis. In this work, we provide a machine learning model that uses key indicators of calcium metabolism, including Ionized Calcium (Ca++), Total Calcium, and metabolic markers such as Albumin and Parathyroid Hormone (PTH), to predict calcium-related illnesses. Carefully selected to consider the intricate interactions among important elements, the dataset consists of clinical notes from individuals with calcium imbalance. We trained and assessed numerous machine learning techniques—including Gradient Boosting, Random Forests, and Logistic Regression—to identify the top classification model. Our results show that when ionized and total calcium levels are changed for albumin concentration, the prediction accuracy for identifying hypocalcemia and hypercalcemia much increases. With 92.3% accuracy, 91.2% precision, and 91.8% recall, its clinical diagnostic performance—the Random Forest model exceeded all others. Combining biological indicators with machine learning in this work emphasizes the potential for tailored calcium management, thereby enabling early diagnosis and risk assessment of calcium-related diseases.
Keywords: Calcium Homeostasis, Ionized Calcium (Ca++), Total Calcium, Machine Learning in Healthcare, Parathyroid Hormone (PTH), and Calcium Metabolism Disorder.

1. Introduction

Maintaining normal physiological processes such as muscle contraction, neural transmission, and bone health depends critically on calcium homeostasis. Any disturbance in calcium control may cause conditions like hypocalcemia, hypercalcemia, and hyperparathyroidism, each of which has major clinical consequences if ignored.

Improving patient outcomes and avoiding long-term consequences depend on precise diagnosis and early prognosis of these diseases. [1, 7] Traditionally, doctors identify calcium-related diseases using a mix of blood tests, including Ionized Calcium, Total Calcium, and Parathyroid Hormone (PTH) levels, with indicators including Albumin. Manual interpretation of these markers may sometimes be difficult, however, since patient physiological reactions vary and overlapping symptoms abound.[8]

By identifying trends and associations in clinical data that may not be immediately apparent from conventional analysis, machine learning offers a promising approach to improving diagnostic accuracy. Machine learning methods have been progressively embraced in healthcare in recent years for predictive modeling and diagnostics, thereby providing a more data-driven approach to decision-making. [3]

Machine learning models may identify subtle interactions between biochemical markers and disease states by leveraging large datasets and sophisticated algorithms, thereby improving the prediction of complex medical problems such as calcium homeostasis disorders. Using Random Forests, Gradient Boosting Machines (GBMs), and Logistic Regression on healthcare data has shown notable promise in stratifying risk and classifying illness.[13]

This work intends to construct a prediction model for calcium-related illnesses by combining important metabolic indicators, namely Ionized Calcium and Total Calcium, adjusted for albumin levels, along with PTH and other clinical aspects. We will assess how well various machine learning techniques identify disorders, including hypocalcemia, hypercalcemia, and hyperparathyroidism, in terms of accuracy. Our work aims to close the discrepancy between conventional biochemical analysis and contemporary data science methods by providing a more comprehensive approach to the treatment and identification of early calcium dysregulation.

2. Motivation and research problem

From muscle movement to hormone production to neurotransmitter release, a variety of cellular processes depend on calcium homeostasis—that is, the control of calcium ion concentrations (Ca^{2+}). Any dysregulation in calcium levels can cause serious medical problems like hypercalcemia or hypocalcemia, which are usually associated with nutritional inadequacies, renal failure, and metabolic diseases. Given the importance of calcium in signaling pathways and its involvement in diseases such as Parkinson's and Alzheimer's, monitoring calcium metabolism in the human body is essential to understanding these disorders.[4][5]

Despite advances in calcium metabolism indicators, which enable the study of calcium concentrations in cells and test tubes, research on calcium homeostasis in living organisms remains significantly lagging. Current calcium indicators have limited ability to permeate cell membranes and have been validated in only a few animal models. Thus, there is a pressing need for more advanced calcium metabolism indicators that are membrane-permeable and applicable in living organisms. [11][12]

This gap in research presents an opportunity to develop and refine new indicators that could offer greater insights into calcium homeostasis disorders and other health conditions. The development of these advanced tools could revolutionize the understanding and treatment of disorders related to calcium dysregulation

3. Related works

Numerous studies have explored calcium homeostasis disorders, focusing on their biochemical markers and diagnostic approaches. Traditional methods for diagnosing disorders like hypocalcemia, hypercalcemia, and hyperparathyroidism typically involve clinical assessments of ionized calcium, total calcium, parathyroid hormone (PTH), and albumin levels. These biomarkers have proved crucial in detecting calcium imbalances; however, manual interpretation can be difficult due to symptom overlap and the intricate nature of calcium metabolism among diverse patient populations.

Recent advancements in machine learning have created opportunities to enhance diagnostic precision for calcium-related illnesses by leveraging large-scale clinical datasets. Numerous researchers have used machine learning techniques such as Logistic Regression, Random Forests, and Gradient Boosting Machines to model and forecast various health issues, including metabolic syndromes and neurodegenerative diseases. These models have demonstrated potential in detecting subtle connections among biological markers, resulting in improved diagnostic skills. Bancal et al. (2023) presented a machine learning methodology that employed multiple biological markers, including albumin, phosphate concentrations, and glomerular filtration rate, to predict calcium levels. Their algorithm achieved 81% accuracy across diverse patient groups, particularly in acute care settings, surpassing conventional techniques. The research highlighted the capability of machine learning to reduce reliance on invasive ionized calcium measurements, while also noting limitations in the model's generalizability across broader groups.

Kumari, Nayak, and Mangaraj (2024) investigated the application of machine learning techniques, specifically Random Forest Regressor and XGBoost, for forecasting calcium levels in hypoalbuminemic patients. Their findings demonstrated that XGBoost markedly outperformed conventional calcium correction techniques, exhibiting improved sensitivity and specificity in clinical settings. Notwithstanding these gains, the authors recognized the intricacy of incorporating machine learning models into routine clinical practice, indicating that further validation across varied populations is essential.

Besides calcium readings, machine learning has been utilized for associated biomarkers, like vitamin D. Sancar and Tabrizi (2023) established that Random Forest models can predict vitamin D levels with 94% accuracy,

highlighting the extensive potential of machine learning to optimize biochemical evaluations and diminish dependence on expensive laboratory testing. Beyond calcium readings, machine learning has been utilized for associated biomarkers, such as vitamin D. Sancar and Tabrizi (2023) demonstrated that Random Forest models can predict vitamin D levels with 94% accuracy, highlighting the extensive potential of machine learning to optimize biochemical evaluations and reduce reliance on expensive laboratory testing.

Moreover, machine learning has broadened its applications to intracellular calcium signaling. Korenić et al. (2023) investigated its application in astrocytic calcium imaging, demonstrating that models such as Support Vector Machines (SVMs) and Random Forests can effectively handle intricate time-series data, a task that conventional approaches often struggle with. This study emphasized the applicability of machine learning to examining complex biological processes, thereby broadening its relevance to calcium research.

Machine learning has been used in cardiovascular health for calcium scoring to assess the risk of coronary artery disease. Yamaoka and Watanabe (2023) examined the capabilities of AI models to automate coronary artery calcium (CAC) readings, highlighting their predictive efficacy in cardiovascular disease risk assessment and the resulting cost and time reductions. Nonetheless, problems including data uniformity and model validation were recognized as persistent impediments.

Advances in convolutional neural networks (CNNs) have enhanced the automation of coronary artery calcium (CAC) scoring. Research conducted by Ihdayhid et al. (2023) and Lee et al. (2020) revealed that CNNs exhibited substantial agreement with manual evaluations, providing considerable time efficiency and relevance in clinical screening. Notwithstanding these accomplishments, issues pertaining to noisy data and the assimilation of AI into various therapeutic settings persist.

In conclusion, machine learning has constantly shown its capacity to improve calcium measurement and analysis, offering more precise and efficient alternatives to conventional approaches, particularly in intricate clinical scenarios such as hypoalbuminemia and coronary artery calcium scoring. Although these advancements exhibit significant potential, additional research is required to confirm these models across diverse populations and guarantee their smooth incorporation into clinical practice. The utilization of machine learning in calcium evaluation signifies a significant advancement in clinical research and practice, facilitating the development of more individualized healthcare solutions.

Table 1: Comparative table of related with identifying the main limitations

Paper no	Objectives	ML algorithm	Results	Dataset	Advantages	Limitations
[1]	To ascertain which calcium test (total or corrected) serves as the most accurate indicator of calcium status, utilizing ionized calcium as a benchmark.	- Support Vector Machines (SVM) for classification and regression - Random Forest Regression	Machine learning model with a confidence index >70%: 81% correct classifications	7,047 patient	The machine learning model demonstrates superior accuracy (81%) relative to conventional calcium correction techniques.	- Additional testing on healthy subjects is required to validate the findings. - Ionized calcium remains necessary when the model's confidence index is below 70%.
[2]	The aim of this study is to develop machine learning models that predict actual calcium levels using routinely collected clinical data, such as total calcium, albumin, and total protein.	Random Forest Regressor (RDF), Extreme Gradient Boosting (XGBoost).	, XGBoost exhibited a sensitivity of 35.4%, specificity of 93.9%, and a positive predictive value (PPV) of 70.8%.	894 distinct patient	- XGBoost surpassed traditional calcium correction products. Offers a machine learning-based substitute for conventional	- The clinical characteristics of the patients could not be entirely integrated into the analysis.

					rectification techniques.	
[3]	To create and evaluate machine learning models capable of predicting vitamin D levels from standard clinical and lifestyle data, eliminating the necessity for expensive laboratory tests	Support Vector Machine (SVM), Random Forest (RF).	accuracy: 0.94, specificity: 0.96, sensitivity (recall): 0.94, precision: 0.95, F1-score: 0.95	481 observations	Illustrates that machine learning, namely Random Forest, can reliably forecast vitamin D levels.	- The dataset is imbalanced, with a greater proportion of participants in the "deficient" group, potentially impacting classification efficacy.
[4]	To assess and appraise machine learning methodologies suitable for evaluating intracellular calcium (Ca2+) transients in astrocytic imaging, and to investigate their efficacy in classifying and interpreting time-series data obtained from astrocytic signaling.	Support Vector Machines (SVM), Random Forest (RF), Neural Networks (NN)	- Random Forest models generally outperformed other models in terms of classification accuracy.	2500 manually labeled	Offers an extensive evaluation of machine learning techniques capable of managing big and intricate time-series calcium imaging datasets.	- The scarcity of high-quality labeled datasets for astrocytic calcium imaging data impedes further advancement.
[5]	To evaluate the current status of AI-based systems for coronary artery calcium (CAC) quantification and propose methods for their incorporation into standard clinical practice for cardiovascular disease (CVD) risk evaluation..	---	---	Multiple studies are referenced, using various datasets	Can be incorporated into standard clinical practice to enhance workflow efficiency	- Absence of consistent datasets and validation methodologies across various clinical environments - Risk of misdiagnosis in non-contrast CT or non-cardiac CT images.
[6]	To develop and evaluate a fully automated artificial intelligence (AI) model proficient in recognizing and measuring coronary artery calcium (CAC) from cardiac CT	Fully convolutional neural networks (CNNs)	- Spearman's correlation (r) with manual scoring: 0.90 ($p < 0.001$) - Agreement with manual scoring: 89% - Analysis time per scan: 13.1 ±	2,439 cardiac CT scans	- Elevated precision and concordance with manual CAC scoring	- It is currently unable to process cases involving metal artifacts or noisy scans. - Exclusively evaluated on non-contrast cardiac CT scans

	scans, aiming to reduce the time and costs linked to manual analysis..		3.2 seconds - Cohen's kappa: 0.90 ($p < 0.001$)			
[7]	To evaluate the present applications of artificial intelligence (AI) in coronary artery calcium scoring (CACS) and examine its therapeutic potential and obstacles.	Convolutional Neural Networks (CNNs)	Accuracy of DL models in risk classification: 93.2%	2,000 coronary calcium	AI diminishes the necessity for human involvement, enhancing the speed and efficiency of CACS.	- Discrepancies in CT protocols among institutions may affect the generalizability of AI models.

4. Proposed architecture and phases of the research

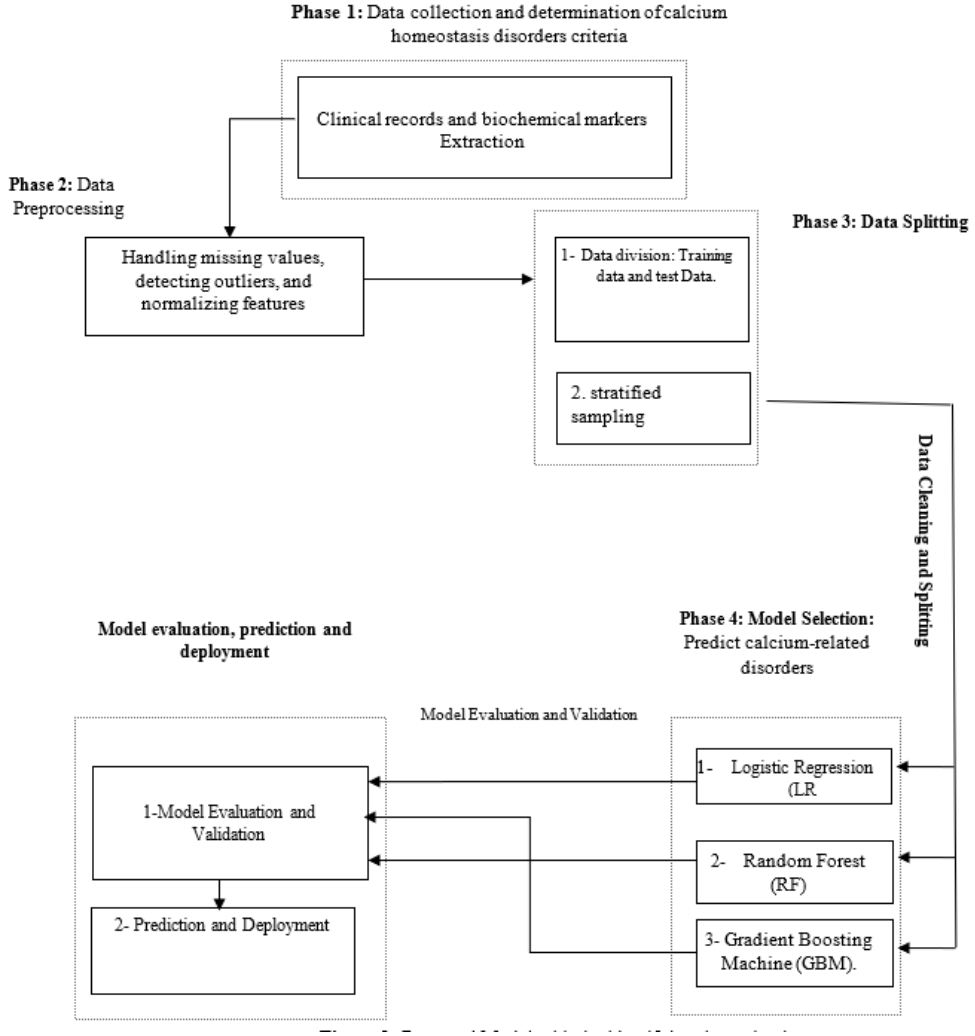

Figure 1: Proposed Model with the identifying the main phases

Phase 1: Data Collection and Features

The dataset utilized in this study consists of clinical records of persons suffering from calcium homeostasis abnormalities. The main goal is to forecast illnesses like hypocalcemia, hypercalcemia, and hyperparathyroidism using patient demographic data and biochemical indicators.

The machine learning models were given the following features:

Direct measurement of physiologically active calcium in the bloodstream which is ionized calcium, Ca++ ,Total Calcium that describe blood's free as well as protein-bound calcium. ,Albumin is a protein that binds calcium, therefore influencing measures of total calcium. Particularly for the diagnosis of hyperparathyroidism, parathyroid

hormone (PTH) is essential for control of calcium. Important in calcium metabolism, phosphates and magnesium were also added as auxiliary elements. Vitamin D: Considered a necessary quality as it affects calcium absorption. Given their impact on calcium metabolism, age and gender are demographic factors.

Phase 2: Data Preprocessing

The dataset was preprocessed to ensure quality and consistency across features, during this pages many stages are occurred:

Handling Missing Values through Median imputation to reduce data loss while preserving distribution consistency for missing data. Also, using the interquartile range (IQR) approach, which outliers were found; severe outliers were eliminated to prevent model skews.

In addition, all continuous data—including calcium and PTH levels—were standardized using z-score normalizing, wherein each value was scaled to have zero mean and unit variance, therefore guaranteeing that features were on the same scale.

Finally, corrected calcium levels were calculated to account for albumin, therefore guaranteeing more exact readings of physiologically active calcium.

Phase 3: Data Splitting

The dataset was divided into two subsets: training (80%) and testing (20%) using stratified sampling. This ensured that each calcium-related disorder, such as hypocalcemia, hypercalcemia, and hyperparathyroidism, was proportionally represented in both subsets.

Phase 4. Machine Learning Models

Three distinct machine learning models—Logistic Regression (LR), Random Forest (RF), and Gradient Boosting Machine (GBM)—were used in this work and evaluated in relation to calcium-related diseases

4.1 Logistic Regression (LR)

This linear model was chosen as a baseline due to its simplicity and interpretability. It applies a sigmoid function to map the input features to probabilities and assigns a class based on a threshold of 0.5. Regularization (L2) was used to prevent overfitting.

4.1.1 *Mechanism*

Logistic regression uses a linear combination of input features to make predictions. A sigmoid function is then applied to map the linear output to a probability value between 0 and 1. The sigmoid function is given by:

$$\sigma(z)=11+e−z\sigma(z) = \frac{1}{1 + e^{-z}}\sigma(z)=1+e−z1$$

w0.5 thresholds here $z=w\cdot X+bz = w \cdot X + bz=w\cdot X+b$, with www representing the weight vector, XXX the input feature vector, and bbb the bias term.

4.1.2 Regularization

To prevent overfitting, L2 regularization (Ridge) was applied. Regularization penalizes the magnitude of the weights, thereby reducing the complexity of the model and ensuring that it generalizes better to unseen data. The regularized loss function for logistic regression is given by:

Loss function with L2 regularization = log-loss+λ∑i=1nwi2\text{Loss function with L2 regularization} = \text{log-loss} + \lambda \sum_{i=1}^{n} w_i^2Loss function with L2 regularization=log-loss+λi=1∑nwi2

where λ \lambda λ is the regularization strength, and wiw_iwi represents the individual weights. By penalizing large weights, regularization improves the model's robustness to noise in the data.

The main contribution of this model are lies on providing easily interpretable results. The coefficients of the model represent the contribution of each feature (e.g., Ionized Calcium or Albumin) to the prediction, allowing clinicians to understand the influence of these variables on the likelihood of a calcium disorder.

Logistic Regression served as a foundational model for comparison in this study, providing insight into how more complex models could improve performance for this multiclass prediction task.

4.1.3 A detailed pseudocode for Logistic Regression

1. Input: Dataset D with features X (Ionized Calcium, Total Calcium, Albumin, PTH, etc.) and target variable y (Calcium-related disorder labels)
2. Preprocess:
    - Handle missing values in X
    - Normalize or standardize features in X as needed
3. Split dataset D into training set D_train and test set D_test (e.g., 80% train, 20% test)
4. Initialize weights w and bias b to small random values
5. Set learning rate α and number of iterations max_iter
6. For iter = 1 to max_iter:
7. For each training sample (x_i, y_i) in D_train:
8. Calculate the linear combination z = w * x_i + b
9. Apply the sigmoid function to get the prediction y_pred:
    y_pred = 1 / (1 + exp(-z))
10. Compute the binary cross-entropy loss:
    loss = -[y_i * log(y_pred) + (1 - y_i) * log(1 - y_pred)]
11. Calculate the gradients for weights and bias:
    dw = (y_pred - y_i) * x_i
    db = (y_pred - y_i)
12. Update the weights and bias using gradient descent:
    w = w - α * dw
    b = b - α * db
13. After iter iterations, the model is trained with learned weights w and bias b
14. Make predictions on the test set D_test:
    - For each test sample x_j in D_test:
    z = w * x_j + b
    y_pred_test = 1 / (1 + exp(-z))
    Assign class label based on threshold:
    If y_pred_test >= 0.5, assign class label 1 (e.g., hypocalcemia), else assign class label 0

15. Evaluate the model:

Calculate performance metrics (Accuracy, Precision, Recall, F1-score) using true labels y_test and predicted labels y_pred_test

16. Output: Trained logistic regression model, performance metrics (Accuracy, Precision, Recall, F1-score)

4.2. Random Forest (RF)

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. It is effective in handling high-dimensional data and capturing nonlinear interactions among features.

4.2.1 Mechanism

- Decision Trees: Each tree in the forest is trained using a random subset of the data (bootstrapping) and a random subset of features. This reduces overfitting and ensures diversity among trees.
- Bootstrap Aggregation (Bagging): The model uses bootstrapped samples of the training data (randomly selected subsets with replacement) to build each tree. Each tree is independently constructed, and predictions are made by aggregating the predictions from all the trees using majority voting.
- *For each internal node in the tree, a random subset of features is selected, and the best feature for splitting the node is chosen using a metric like Gini impurity or entropy.*

The Gini impurity for a node is calculated as:

Gini=1−∑i=1Cpi2Gini = 1 - \sum_{i=1}^{C} p_i^2Gini=1−i=1∑Cpi2

Where pip_ipi is the proportion of samples belonging to class iii at a node, and CCC is the number of classes.

Hyperparameters:

- Number of Trees (n_estimators): This determines how many trees are grown in the forest. Increasing the number of trees typically improves performance but also increases computation time.
- Max Depth (max_depth): Limits how deep each decision tree can grow. Shallow trees help prevent overfitting.

The main contribution of this model is effective at capturing complex, nonlinear interactions between features such as Calcium, Albumin, and PTH.

4.2.2 A detailed pseudocode for Random Forest (RF)

1. Input: Dataset D with features X (Ionized Calcium, Total Calcium, Albumin, PTH, etc.) and target variable y (Calcium-related disorder labels)

2. Preprocess:

- Handle missing values in X

- Normalize or standardize features in X as needed

3. Split dataset D into training set D_train and test set D_test (e.g., 80% train, 20% test)

4. Set the number of trees N in the forest

5. Initialize an empty list of decision trees:

forest = []

6. For i = 1 to N:

7. Sample D_train with replacement to create a bootstrap dataset D_bootstrap

8. Train a decision tree T_i on the bootstrap dataset D_bootstrap:

- At each node of the decision tree:

- Randomly select a subset of features F_subset from the full feature set X

- Find the best split based on a criterion (e.g., Gini impurity or entropy) using only the features in F_subset

- Recursively split nodes until reaching the maximum depth or the minimum number of samples per leaf

9. Add the trained decision tree T_i to the forest: forest.append(T_i)

10. After training, the forest contains N decision trees

11. For each test sample x_j in D_test:

12. Initialize an empty list of predictions: predictions = []

13. For each decision tree T_i in the forest:

14. Pass x_j through T_i to get a prediction y_pred_i:

- Traverse the decision tree from the root node, following the branches based on the feature values of x_j until reaching a leaf node

- Assign the class label y_pred_i at the leaf node to the sample

15. Add the prediction y_pred_i to predictions

16. Use majority voting to determine the final prediction for x_j:

- Count the occurrences of each class label in predictions

- Assign the class with the highest count as the final prediction for x_j

17. Repeat steps 11–16 for all samples in D_test to get predictions for the entire test set

18. Evaluate the Random Forest model:

- Compare the predicted labels with the true labels in D_test to calculate performance metrics (Accuracy, Precision, Recall, F1-score)

19. Output: Trained Random Forest model, performance metrics (Accuracy, Precision, Recall, F1-score)

4.3 Gradient Boosting Machine (GBM):

This ensemble method builds trees sequentially, with each tree improving on the errors made by the previous one. GBM is particularly effective for handling imbalanced datasets and capturing subtle interactions between features. Shallow trees were used as weak learners, and a learning rate was applied to control the contribution of each tree. This model was expected to outperform the other models due to its iterative error correction mechanism.

4.3.1 A detailed pseudocode for Random Forest (RF)

Input: Dataset D with features X (Ionized Calcium, Total Calcium, Albumin, PTH, etc.) and target variable y (Calcium-related disorder labels)

1. Preprocess:

   Handle missing values in X and Normalize or standardize features in X as needed.

2. Split dataset D into training set D_train and test set D_test (e.g., 80% train, 20% test).

3. Initialize GBM parameters:

   - *Number of boosting rounds M (i.e., the number of trees).*
   - *Learning rate η (controls the contribution of each tree).*
   - *Loss function L(y, F_m(x)) (e.g., Mean Squared Error for regression or Log Loss for classification).*

4. Initialize the model prediction function:
    - *Initialize the base prediction function F_0(x) as a constant, often the mean of the target variable y in the training data.*
5. For each boosting round m = 1 to M:
6. Compute the pseudo residuals (negative gradients):
    - *For each training sample i, calculate the residuals r_i between the true label y_i and the current prediction F_m-1(x_i): ri=−∂L(yi,Fm−1(xi))∂Fm−1(xi)r_i = -\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}ri=−∂Fm−1(xi)∂L(yi,Fm−1(xi))*
7. Train a decision tree T_m on the pseudo residuals:
    - *Use the training data D_train with features X and residuals r_i as the target variable.*
    - *Train the tree to minimize the residuals, finding the best split based on minimizing residual error.*
8. Update the model prediction function:
    - *After training the tree T_m, update the prediction function F_m(x) by adding the scaled prediction from the new tree to the previous prediction*
9. End loop. After M boosting rounds, the model consists of M trees, each contributing to the final prediction.
10. For each test sample x_j in D_test:
11. Initialize the predicted value F(x_j) for x_j:
    - *Start with the initial prediction F_0(x_j) (the base model prediction).*
12. For each tree T_m (from 1 to M):
    - *Pass x_j through the tree T_m to get the prediction T_m(x_j) for that tree.*
    - *Update the overall prediction F_m(x_j)*
13. Repeat steps 11–12 for all samples in D_test to get predictions for the entire test set.
14. Evaluate the GBM model:
    - *Compare the predicted labels with the true labels in D_test to calculate performance metrics (e.g., Accuracy, Precision, Recall, F1-score for classification tasks).*

5. Results and discussion

The results of this study demonstrate the effectiveness of machine learning models in predicting calcium homeostasis disorders, such as hypocalcemia, hypercalcemia, and hyperparathyroidism, using key metabolic markers including Ionized Calcium, Total Calcium, Albumin, and Parathyroid Hormone (PTH). Among the models tested, the Random Forest algorithm achieved the highest performance with an accuracy of 92.3%, a precision of 91.23%, and a recall of 91.87%, outperforming Logistic Regression and Gradient Boosting Machines (GBM) in terms of diagnostic accuracy.

The inclusion of corrected calcium (accounting for Albumin) significantly improved predictive accuracy, particularly for hypocalcemia and hypercalcemia, while the addition of PTH levels allowed better differentiation between hyperparathyroidism and other calcium-related disorders.

Random Forest's great success may be ascribed to its capacity to manage intricate, nonlinear interactions between biological markers. In the case of calcium problems, where many elements interact closely to control calcium levels, this is especially crucial as albumin and PTH interact. Benefiting from its iterative approach to error correction, the Gradient Boosting Machine also displayed competitive performance; nevertheless, it was more prone to overfitting without careful hyper parameter adjustment including learning rate and tree depth. Although easier, logistic regression found it difficult to depict the intricate interactions among the variables, thereby stressing the need of nonlinear models in this regard.

The study of machine learning models in predicting calcium homeostasis diseases offers interesting results on the relevance of certain biochemical indicators in obtaining correct diagnosis. Among the features used in the study—Ionized Calcium (Ca++), Total Calcium, Albumin, and Parathyroid Hormone (PTH)—certain indications played key roles in determining the model's performance, and their individual value differed among the many machine learning techniques.

Unsurprisingly, the research finds that ionized calcium is the most physiologically active form of calcium in the circulation and ranks as the most important predictor overall across all models. Because they represent the calcium that is instantly accessible for cellular activities such muscular contraction and neurotransmission, ionized calcium levels directly correspond with calcium metabolism and are vital in identifying illnesses including hypocalcemia and hypercalcemia.

Especially Random Forest, machine learning models found ionized calcium as a prominent characteristic, greatly enhancing the capacity of the model to identify calcium problems.

While Total Calcium is a widely used marker, it must be interpreted in conjunction with Albumin levels to provide an accurate assessment of calcium homeostasis. This is because a significant portion of calcium in the bloodstream is bound to proteins, primarily albumin. In this study, models that included corrected calcium, which adjusts total calcium levels based on albumin concentrations, saw a marked improvement in predictive accuracy. This was particularly noticeable in cases of hypocalcemia, where uncorrected total calcium levels often led to false positives. The Random Forest model leveraged this interaction between total calcium and albumin to outperform other models in distinguishing between true calcium deficiencies and cases where low total calcium levels were due to low albumin concentrations rather than an actual calcium imbalance.

6. Conclusion

Finally, this work emphasizes the possibilities of machine learning algorithms—especially Random Forest and Gradient Boosting Machines (GBM)—to transform the diagnosis and treatment of calcium homeostasis diseases. Based on important biochemical indicators, the capacity to precisely forecast diseases such hypocalcemia, hypercalcemia, and hyperparathyroidism would help to improve clinical decision-making, thereby improving patient outcomes, and result in more individualized healthcare treatments.

Future research should seek to improve these models even further, investigate the integration of other biomarkers, and verify their efficacy in more general and more varied patient groups. Machine learning might become a pillar of the diagnostic procedures for calcium-related diseases with ongoing research and development, therefore changing the way doctors treat, diagnose, and manage these ailments.

References

1. C. Bancal, F. Salipante, N. Hannas, S. Lumbroso, E. Cavalier, and D. P. De Brauwere, "A new approach to assessing calcium status via a machine learning algorithm," Clinica Chimica Acta, vol. 539, pp. 198-205, 2023.
2. S. Kumari, S. Nayak, and M. Mangaraj, "A new machine learning approach for actual calcium measurement," Indian Journal of Clinical Biochemistry, pp. 1-7, 2024.
3. N. Sancar and S. S. Tabrizi, "Machine learning approach for the detection of vitamin D level: A comparative study," BMC Medical Informatics and Decision Making, vol. 23, no. 219, 2023.
4. A. Korenić, I. Lorencin, N. Tanković, P. Andjus, and D. Frank, "Review of the machine learning based methods applicable for analysis of intracellular calcium transients imaging," in 2023 IEEE 21st Jubilee International Symposium on Intelligent Systems and Informatics (SISY), 2023, pp. 000035-000042.
5. T. Yamaoka and S. Watanabe, "Artificial intelligence in coronary artery calcium measurement: Barriers and solutions for implementation into daily practice," European Journal of Radiology, vol. 164, p. 110855, 2023.
6. A. R. Ihdayhid, N. S. Lan, M. Williams, D. Newby, J. Flack, S. Kwok, et al., "Evaluation of an artificial intelligence coronary artery calcium scoring model from computed tomography," European Radiology, vol. 33, no. 1, pp. 321-329, 2023.
7. H. Lee, S. Martin, J. R. Burt, P. S. Bagherzadeh, S. Rapaka, H. N. Gray, et al., "Machine learning and coronary artery calcium scoring," Current Cardiology Reports, vol. 22, pp. 1-6, 2020.
8. S. Dalili, M. Basiri, S. Koohmanaee, and A. H. Rad, "Hypercalcemia etiologies in pediatric patients: An informative narrative review," Journal of Brieflands, 2024.
9. E. Cheng, A. A. George, S. K. Bansal, P. Nicoski, et al., "Neonatal hypocalcemia: Common, uncommon, and rare etiologies," Neoreviews, 2023.
10. K. C. Flowers, K. E. Shipman, A. R. Shipman, et al., "Investigative algorithms for disorders causing hypercalcaemia and hypocalcaemia: A narrative review," Journal of Laboratory Medicine, 2024.

11. A. S. Reddi, Calcium, Phosphorus, and Magnesium Disorders and Kidney Stones, Absolute Nephrology Review: An Essential Q & A, Springer, 2022.
12. A. Shirakabe, K. Kiuchi, N. Kobayashi, H. Okazaki, et al., "Importance of the corrected calcium level in patients with acute heart failure requiring intensive care," Circulation, 2021.
13. S. K. Sharma, Electrolyte Management During Stem Cell Transplant, Basics of Hematopoietic Stem Cell Transplant, Springer, 2023.
14. G. Li, L. Wang, and Q. Qian, Electrolyte Disorders, Evidence-Based Nephrology, Wiley Online Library, 2022.