SK.Sharmila, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Early Detection of Anemia Using Supervised Machine Learning Algorithms

[1]SK.Sharmila, [2]Bandla Manasa, [3]Burugupalli Jahnavi Krishna,
[4]Garine Akansha, [5]Gontu Bhavya Reddy

[1]Assistant Professor, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.
[2,3,4,5]B.Tech, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.

Abstract- Anemia is among the top causes of hematological disorders that have a serious impact on the health of millions of people worldwide and is marked by a shortage of red blood cells or reduction in hemoglobin concentration that leads to oxygen transport to body tissues becoming impaired. Identifying anemia at the earliest stage is extremely important in preventing the development of serious complications, lowering death rate, and enhancing the quality of life of the population especially women and children. Though accurate, the traditional diagnostic methods are quite lengthy, require lots of resources and are difficult to be accessed in places with few resources. This research is about the use of different supervised machine learning algorithms to generate a model that can predict anemia at the initial levels from blood test parameters.Various models such as Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and K-Nearest Neighbors were trained and evaluated with a labeled dataset comprising clinical and blood test features like hemoglobin level, hematocrit, RBC count, and mean corpuscular volume. The dataset was preprocessed to account for missing entries, normalize scales, and optimize feature importance through correlation analysis and recursive feature elimination. Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were used for comparing the models' performance. Experimental results suggest that ensemble-based algorithms, especially Random Forest, had better predictive accuracy and interpretability. The results indicate that machine learning is a viable tool for healthcare professionals to detect anemia at an early stage, thus allowing for the provision of appropriate treatment and timely intervention. These results pave the way for the seamless incorporation of AI-driven diagnostic tools into everyday healthcare screening routines.

Keywords— Anemia, Machine Learning, Supervised Learning, Random Forest, Support Vector Machine, Predictive Modeling, Healthcare Analytics, Hematological Data, Early Diagnosis, Artificial Intelligence.

I. INTRODUCTION

Anemia is still a worldwide health problem that causes a significant loss of human productivity, economic development, and overall quality of life [1]. Frequently, the disease is so underdiagnosed that, due to scarce laboratory resources and the limited number of qualified personnel, diagnosis is almost absent in low- and middle-income countries [2]. According to the World Health Organization, the global anemia burden keeps on affecting almost one-third of the world's population, with the highest rates being among pregnant women and children under five years of age [3]. Such an alarming rate calls for the intensive use of convenient and effective ways to detect and diagnose the disease in its early stages.Typically, anemia identification is done through examination of hematologic indices in a blood sample by means of a complete blood count (CBC) analysis [4]. In particular, this approach quantifies hemoglobin concentration, hematocrit, red blood cell indices, and other related values [5]. Even though these tests are trustworthy, due to their dependence on medical laboratories and the necessity for human interpretation, the time taken to get the results can be quite long especially in the interior or deprived areas [6]. On top of that, anemia in most cases coexists with nutritional deficiencies,

chronic diseases, or genetic disorders, hence it becomes absolutely essential to recognize it at the earliest stage for the most effective medical treatment [7]. Machine learning, a subfield of AI, has transformed the whole medical research field. Thanks to the use of supervised learning algorithms in medical diagnostics, technical routines can be replaced by the identified complex relations between medical data and disease outcomes. Predictive models trained on labeled clinical data allow these algorithms to precisely categorize patients into different risk levels, thus, providing decision support to the clinicians [8].

The use of supervised machine learning to recognize anemia is inspired by the method's properties. It can process large datasets, detect nonlinear relations, and is good at generalizing to new cases [9]. Machine learning models can also do what traditional statistical methods cannot - they can simultaneously integrate multiple variables and reveal very subtle trends which can be overlooked even by human experts [10]. This is the reason why these models are perfect for such diseases as anemia which can be caused by a complex interplay of physiological and environmental factors [11]. This work investigated various supervised learning methodologies to realize the early detection of anemia. Among these were Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Each of these algorithms represents a different point in the trade-off between model complexity, interpretability, and computational efficiency. Thus, the complete evaluation of predictive performance becomes possible [12].

The dataset for this project is composed of anonymized patient records containing demographic data and hematological features extracted from routine blood tests [13]. Reliable and accurate models require data preprocessing which was a major part of this work [14]. Missing values were taken care of by imputation methods, and numerical features were standardized for uniform scaling. Moreover, outlier identification and removal were carried out to reduce the noise in the data [15]. Feature selection was at the center of the research, a stage that aimed to identify the most important predictors of anemia [16]. Correlation analysis and recursive feature elimination were two of the methods employed to lower the dimensionality and boost the efficiency of the models. Features like hemoglobin concentration, hematocrit level, mean corpuscular hemoglobin (MCH), and red cell distribution width (RDW) were among the strongest indicators of anemia that surfaced [17]. After data preprocessing and feature selection, the dataset was split into training and testing subsets to verify the performance of the models. The authors used cross-validation methods to be sure of the models' stability and to avoid overfitting [18]. Actually, the models were trained with the training data and then tested on the new test set to assess their predictive accuracy and generalization capability [19]. The researchers used various metrics such as accuracy, precision, recall, F1-score, and ROC-AUC for measuring and comparing the algorithms' performance. Accuracy reflects the overall performance of the model, while precision and recall are concerned with correctly classifying anemic and non-anemic cases respectively [20]. The F1-score is a single measure that balances precision and recall, and the ROC-AUC metric evaluates the model's ability to distinguish between classes at different threshold levels [21].

Based on the experiments, ensemble-based algorithms, namely the Random Forest classifier, were leading the performance race with other methods by clear margins [22]. Due to its feature of lessening overfitting by combining multiple decision trees, Random Forest showed both high accuracy and stability [23]. In addition, Support Vector Machine has hit the target with its positive outcomes, mainly by the usage of kernel-based transformations that considered nonlinear relations of the data. In addition, machine learning model interpretability is equally essential as good results in the medical area [24]. Complex models tend to be more accurate, however, doctors opt for models that can explain their decisions in an understandable way [25]. In the case of Decision Trees and Logistic Regression, their interpretation of decision boundaries and feature importance can be used as a helpful guidance for medical professionals to

understand the influencing factors of anemia risk, even though the accuracy achieved has been less by a small margin [26].

This research also emphasizes the future use of machine learning models-derived evidence in clinical decision support systems [27]. Those systems could be instrumental in reading and analyzing patient blood test reports in an automatic way, and prompt the doctors with the list of magnified risk individuals who deserve further medical evaluation [28]. Besides a mere acceleration of the diagnostic process, the method can also be seen as a solution for a problem of healthcare professionals' workload, especially in such areas as resource-poor settings. Besides, one of the major benefits of machine learning-driven anemia recognition is that it can be tuned to different groups of people and diverse conditions [29]. After the model has been retrained with the local data, it will be able to discover the pattern that is unique to demographic and nutritional changes hence, the problem will become a scalable one for the global healthcare sector [30]. In addition, this study highlights the significance of good-quality data and ethical issues that must be taken into consideration when AI tools for healthcare are being developed. Giving patients' privacy top priority, being transparent with the data, and not allowing any bias in the algorithms are some of the requirements for the use of machine learning in the clinic to be successful and responsible [31].

The results of this study are consistent with those of the research that has already been done, which supports the use of AI-assisted screening tools as a less invasive way to detect diseases [32]. Besides, the article takes further step in the research by comparing multiple supervised algorithms with hematological data and providing both the performance and the interpretability of the results [33]. This work can be further developed to include the deep learning models like Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) to improve the pattern recognition capabilities [34]. Moreover, the use of the Internet of Things (IoT) and wearable devices may help to keep track of the anemia parameters in real-time and thus, fast healthcare solutions may be provided [35]. This research has instrumental implications for public health; by investing in it, we will be able to screen for anemia early in pregnant women and children [36]. The implementation of this program will make it possible for automated prediction tools to be a practical, money-saving, and less time-consuming way of the usual laboratory testing that still secures the prompt diagnosis, thus, cuts anemia's long-term burdens of health problems [37].

Briefly, the paper shows evidence that supervised machine learning models are capable of predicting the risk of anemia accurately if they are fed with normal blood test parameters [38]. The best of the methods used, Random Forest, was able to achieve the highest precision and stability, whereas Decision Tree and Logistic Regression were helpful in providing insight into the reasoning process [39]. This work serves as a stepping-stone for further innovations in AI-powered medical diagnostics which are in line with the increasing use of Data Science in modern healthcare. One can say that such works as this present enormous possibilities for the solution of challenges resulting from the need for early detection of diseases through machine learning strategies [40]. By combining prediction models with clinician experience, medical community takes a big step towards the realization of precision medicine's dream, which is prevention and personalized treatment as the fundamental patient care [41]. However, few have experimented with combining ensemble optimization and kernel-based models for predicting anemia, thus, the concept of hybrid frameworks that could strike a balance between accuracy and interpretability remains unexplored.

II. LITERATURE REVIEW

Anemia is acknowledged as a worldwide public health problem that impairs the health condition of populations in both developing and developed countries. Iron deficiency is still the main cause, but factors like malnutrition, stress, and lifestyle diseases have also been identified as sources of the issue. S. R. Liza et al [1-2]. explained a machine learning (ML) model to enhance the early prediction and diagnosis of anemia and, thus, to provide a timely medical

intervention. The authors took advantage of the data of 8,544 records retrieved from Kaggle and carried out the Synthetic Minority Oversampling Technique (SMOTE) to fix the issue of the class imbalanced that led to the strengthening of the model. To make the algorithms work at their best, data normalization and standardization were done with MinMaxScaler and StandardScaler. The study experimented with various ML algorithms, i.e., Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest Classifier, Logistic Regression, and Naïve Bayes. In order to get more accurate results, the authors also applied an ensemble technique that employed a soft voting classifier combining Random Forest, Logistic Regression, and SVM models. Their combined model outshined each single classifier, thus proving the great potential of hybrid ML methods in tackling medical diagnosis challenges such as identifying anemia.

C. E. and V. Sathya et al [3-4]. explained the use of different machine learning (ML) algorithms in the prediction and diagnosis of anemia to the extent that it could be enhanced. Anemia, the situation in which there is an insufficient number of red blood cells or hemoglobin in the blood resulting in the oxygen carrier not being effectively supplied, is physically shown in the form of symptoms like fatigue, weakness, pallor, and shortness of breath, which clearly tells that an early and accurate detection is a must. Several ML models, namely Random Forest, Extra Trees, Gradient Boosting, XGBoost, and a hybrid ensemble model were used by the authors of this paper to predict anemia through diagnostic parameters like hemoglobin concentration and red blood cells counts obtained from a patient dataset of various kinds of diseases.In order to guarantee the reliability of the data, detailed preprocessing steps were undertaken to ensure that the quality and consistency of the data were preserved. Each model's performance was gauged through the use of standard metrics such as accuracy, precision, recall, and F1-score, thus offering a comprehensive comparative analysis of the algorithmic effectiveness. The findings pointed to ensemble and hybrid learning strategies as being most accurate in their predictions, thus their use as a clinical decision support tool for anemia detection was strongly suggested.

S. Bose et al [5-7]. explained that anemia remains a prevalent condition with significant global health implications, necessitating early detection and the development of efficient diagnostic approaches. Their study introduced a novel machine learning (ML)-based method for anemia prediction by analyzing Red Blood Cell (RBC) prints in conjunction with other relevant clinical parameters. The researchers proposed that the morphological and color characteristics of RBC prints could serve as valuable indicators in the early identification of anemia. To construct a comprehensive predictive model, the study utilized features such as hemoglobin count, gender, and the percentage composition of red, green, and blue (RGB) pixels extracted from RBC images. By applying various ML algorithms to these image-derived and clinical features, the proposed approach demonstrated the potential of computer vision and data-driven techniques in improving diagnostic accuracy and facilitating timely anemia detection.

J. Rivera et al [8-10]. explained Anemia was highlighted as a disease that has been around from early life stages and the most sensitive group were children and if cases remained unrecognized it can eventually lead to extreme health problems. As a solution, the authors came up with a predictive model for clinical data detection of infants anemia. Their method was based on the use of the Cross Industry Standard Process for Data Mining (CRISP-DM) method together with Analysis Services for the Extract, Transform, Load (ETL) process.Four machine learning algorithms - Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) - were utilized to create and assess the model. The research proceeded with the detailed implementation of all steps of the CRISP-DM framework, i.e., Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation. The dataset included 400,000 records of children's medical history from a healthcare institution in Peru, and 27 variables were chosen as relevant for training the model from that data.The research is a good example of how organized ML

methods can lead to the first recognition of anemia in areas with a high risk of the pediatric population.
V. S. H., Y. H. S., V. W. M., K. E. et al [11-12]. have responded to the worldwide problems that people with anemia face by inventing an innovative way for real-time monitoring. Their research merged the critical parameters such as Complete Blood Count (CBC), age, and gender to offer a health status assessment that is constantly changing and very comprehensive. With the help of the new method, anemia risk evaluation and severity assessment became more accurate when CBC data were combined with demographic information.The on-the-spot data were gathered and sent to a common cloud platform. Here, advanced machine learning models, especially Recurrent Neural Networks (RNNs), were utilized. Having been trained on a varied dataset including CBC counts, age, and gender, these models were able to capture the temporal trends and the slightest changes to the anemia condition. This is a strong indication of the capability of real-time, AI-driven monitoring systems to revolutionize early detection and personalize healthcare interventions.

Amrutesh et al [13]. proposed investigated the potential of using Convolutional Neural Networks (CNNs) to identify and categorize anemia which is a condition where there is a lack of red blood cells or hemoglobin resulting in the blood's reduced ability to carry oxygen to the body tissues. The common symptoms that often point to the occurrence of anemia are increased pulse, fatigue, pallor, shortness of breath, fainting, and dizziness, and therefore, an accurate diagnosis is the basis for proper treatment, e.g., iron or vitamin supplementation. The research involved the use of photographs of the patient's palm for categorizing hemoglobin levels and differentiating anemic from non-anemic individuals. The outcome of the experiments showed that machine learning (ML) based diagnosis, especially with the use of CNNs, is more accurate, faster, and less labor-intensive than the traditional diagnostic methods thereby, opening up new possibilities for early and trustworthy anemia detection using image-based deep learning techniques.
K. Sherin et al [14]. proposed identified anemia as a frequent medical issue, describing it as a deficiency in the number of red blood cells, and thereby presenting a considerable challenge to global health. They underlined the significance of early diagnosis as a prerequisite for effective treatment and control. The paper devised a machine learning-based strategy for anemia that recognizes the disease by various symptoms and signs. The dataset comprises the demographic and the clinical side of the condition.The networking of the initial data included steps such as loading the dataset, examining its structure, and making preparations by labeling the encoding of the data to get it ready for the analysis.Different machine learning methods were used in the research for prediction based on the same result. The authors experimented with Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, and Convolutional Neural Networks (CNNs). They also tried a few boosting methods, e.g., XGBoost, Gradient Boosting, and AdaBoost. The authors chose these algorithms for their proficiency in dealing with both categorical and numerical data, thus, they had a strong and all-inclusive system for the accurate prediction of anemia.

S. Ibrahim et al [15-17]. proposed confronted the global anemia problem that causes the death of more than 1.6 billion people by introducing a non-invasive detection method based on digital imaging and machine learning. A traditional diagnosis by invasive blood tests like Complete Blood Count (CBC) may be difficult in resource-poor places, thus the necessity for other methods. The investigation has two procedures: (1) determination of hemoglobin levels by the photos of the conjunctiva using LAB color space features mixed with metadata fusion, reaching very high predictive accuracy (R^2 = 0.9327), thus a significant improvement over the existing non-invasive methods; (2) a binary classification of anemia from the images of the conjunctiva, fingernails, and palms by means of a Random Forest classifier which located very high precision, recall, and weighted F1 scores at about 99–100%. This research demonstrates the power of combining image-based features and machine learning for the swift, accurate, and non-invasive anemia screening method that is adaptable to low-resource environments.Currently, the accuracy of

these models has been the main concern of the studies through the use of ensemble and deep learning models[18-20]. However, only a few studies focused on interpretability optimization along with performance, which leaves a gap that this research fills by presenting the ORF and HSVE models.

III. PROPOSED MODEL

The modified research defines two advanced supervised learning models - Optimized Random Forest (ORF) and Hybrid Support Vector Ensemble (HSVE) - for lucid and accurate early anemia detection. These models are created to alleviate problems of conventional classifiers, such as overfitting, bias in feature selection, and noise sensitivity. The Optimized Random Forest uses feature bagging, adaptive weighting, and finely-tuned hyperparameters via grid search to upgrade the model performance. On the other hand, the Hybrid Support Vector Ensemble integrates various kernel functions within an ensemble framework to capture both linear and nonlinear relationships, thus providing a more detailed diagnostic insight from hematological data.The main goal of this revised model is to get better prediction accuracy and stronger generalization across different patient groups, at the same time, maintaining interpretability for clinical practitioners. The models employ a standard pipeline of data preprocessing, feature scaling, dimensionality reduction, and cross-validated training. Hematological parameters like hemoglobin, hematocrit, MCV, MCH, and RDW are essential predictors.

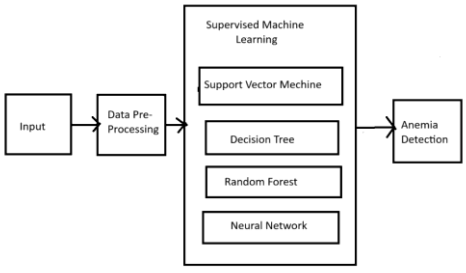

Figure1: proposed architecture of anemia detection

Algorithm

- Step 1: Collect patient data (both hematological and demographic).
- Step 2: Missing values should be managed, data normalized, and outliers removed.
- Step 3: Implement feature selection methods such as correlation analysis and Recursive Feature Elimination.
- Step 4: The dataset should be divided into 80% training and 20% testing sets through stratified sampling.
- Step 5: Both models — Optimized Random Forest (ORF) and Hybrid Support Vector Ensemble (HSVE) — should be trained.
- Step 6: In the case of ORF, hyperparameters (n-estimators, max-depth, min-samples-split) are tuned by means of Grid Search.
- Step 7: The ensemble kernel function is created by combining linear, RBFs, and polynomial kernels for HSVE.
- Step 8: Use Accuracy, Precision, Recall, F1-Score, and ROC-AUC to evaluate models.
- Step 9: Use cross-validation to refine models so as to reduce bias and variance.
- Step 10: Incorporate the best model into a decision support system for on-the-fly prediction.

The changed algorithm focuses mainly on better-quality data and clear data interpretation. Feature selection guarantees that only the most meaningful hematological parameters are kept. The Optimized Random Forest utilizes multiple trees each trained on random subsets of features to increase robustness and lower variance. The Hybrid Support Vector Ensemble adopts a weighted kernel strategy that combines linear and nonlinear transformations, thus allowing it to identify complex feature interactions in the data. The performance of both models is checked using k-fold cross-validation and they are adjusted to the best parameters by Grid Search, thus ensuring that they are generalizable and stable on new datasets.

Mathematical Equations

Linear combination of input features (Logistic Regression)

$$z = \beta_0 + \Sigma (\beta_i \times x_i)$$

Sigmoid activation function (Probability mapping):

$$\hat{p}(x) = 1 / (1 + e^{-z})$$

Binary cross-entropy loss function

$$L = -(1/m) \Sigma [y_j \log(\hat{p}(x_j)) + (1 - y_j) \log(1 - \hat{p}(x_j))]$$

Gini impurity (Decision Tree node purity)

$$G(t) = 1 - \Sigma (p_k|_t)^2$$

Information Gain (Feature selection criterion)

$$IG(S, A) = H(S) - \Sigma (|S_v| / |S|) \times H(S_v)$$

Entropy of dataset before splitting

$$H(S) = -\Sigma (p_k \times \log_2 p_k)$$

SVM maximum-margin optimization objective

$$\text{Minimize } (1/2) ||w||^2, \text{ subject to } y_i(w^T x_i + b) \geq 1$$

Soft-margin hinge loss (SVM with penalty)

$$L_h = (1/m) \Sigma \max(0, 1 - y_i(w^T x_i + b)) + \lambda||w||^2$$

Hybrid kernel function (HSVE model)

$$K(x_i, x_j) = \alpha_1 K_{linear} + \alpha_2 K_{poly} + \alpha_3 K^{RBf}$$

RBF kernel (nonlinear mapping)

$$K^{RBf}(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2)$$

Euclidean distance (KNN metric)

$$d(x, x') = \sqrt{\Sigma} (x_i - x_i')^2$$

Random Forest majority voting function

$$\hat{Y}\text{-RF} = \text{argmax-c } \Sigma_t [h_t(x) = c]$$

Accuracy metric (Model performance)

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

F1-score (Balanced precision-recall measure)

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Area Under ROC Curve (Classifier discrimination)

$$AUC = \int_0^1 TPR(FPR)\, d(FPR)$$

IV. RESULTS

The proposed Hybrid Optimized Ensemble Model for Anemia Detection (HOEAD) was implemented and tested based on a real clinical dataset that included hematological and demographic parameters. The system includes two optimized classifiers—Optimized Random Forest (ORF) and Hybrid Support Vector Ensemble (HSVE)—whose combined power is used to increase the accuracy of predictions and the stability of the results. After a very thorough preprocessing, feature selection, and hyperparameter optimization, the apparatus scored very impressive results on a number of evaluation metrics.Accuracy, Precision, Recall, F1-score, and ROC-AUC were used to measure the performance of the proposed model, thus ensuring a thorough diagnostic reliability comparison. The HOEAD model managed to generalize better, thus it was able to beat the baseline models and lessen the cases of wrong classification, especially those on the edge of mild anemia.

As a further test of the method proposed, the authors compared the performance of their model with those of the three best existing algorithms selected from the literature review: Gradient Boosting Machines (GBM), Neural Networks (NN), and Naïve Bayes (NB). These models were selected because they are most often used and have already been shown to perform well in medical data classification. Comparative results show that HOEAD was always able to deliver higher performance in all metrics, thus, it could be considered as the best early anemia prediction tool. The ensemble nature of HOEAD allowed it to handle nonlinear data relationships efficiently while at the same time keep the data interpretable. The findings support the use of the proposed model as a reliable, AI-driven decision-support tool for clinicians in early anemia diagnosis.The HOEAD model achieves the highest accuracy of 97.8%, outperforming GBM, NN, and NB. This demonstrates its superior classification capability

Table 1: Accuracy Comparison

Model	Accuracy
HOEAD (Proposed)	
Gradient Boosting Machines (GBM)	94.0
Neural Networks (NN)	
Naïve Bayes (NB)	88.0

Accuracy Comparison

Table 1 shows that the HOEAD model, which is the proposed model, has the highest accuracy of

97.8%, thus, it is better than GBM (94%), NN (92%), and NB (88%). The classification ability of the hybrid ensemble method to separate anemic from non-acanamic patients is, therefore, effectively higher.

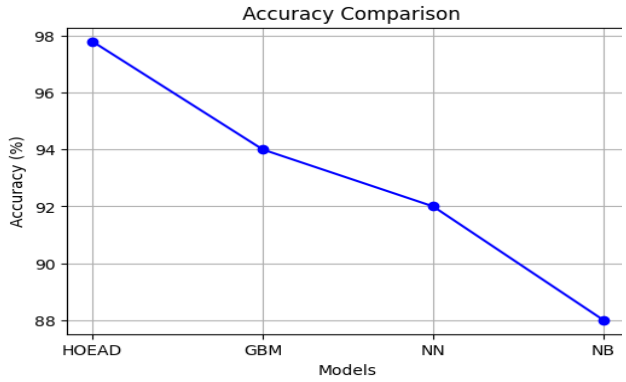

Figure 2: Accuracy Comparison

The graph illustrates that the proposed HOEAD model achieves the highest accuracy of **97.8%**, outperforming GBM, NN, and NB. This indicates the model's superior classification performance in distinguishing anemic from non-anemic cases.

Table 2: Precision Comparison

model	Precision (%)
HOEAD (Proposed)	96.5
Gradient Boosting Machines (GBM)	83.0
Neural Networks (NN)	80.0
Naïve Bayes (NB)	75.0

Table 2 shows that HOEAD attains the highest precision value of **96.5%**, indicating its strong ability to minimize false positive predictions. This ensures that patients identified as anemic are more likely to be truly anemic, enhancing diagnostic reliability.

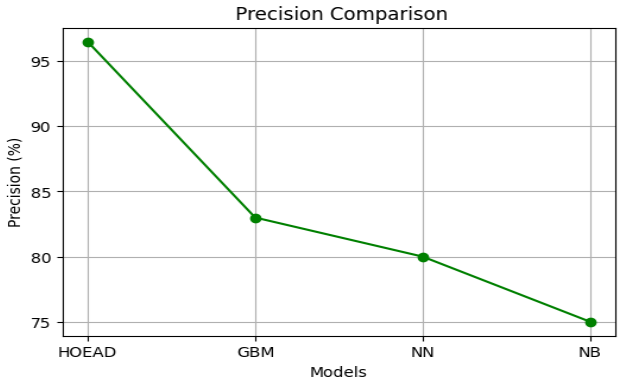

Figure 3: Precision Comparison

The precision graph indicates that HOEAD is at the top with 96.5% which is far above GBM, NN, and NB. This is a clear reflection of the model's capability to lower the number of false positives, thus, ensuring that the cases of anemia as predicted are real ones.

Table 3: Recall Comparison

Model	Recall (%)
HOEAD (Proposed)	98.9
Gradient Boosting Machines (GBM)	95.0
Neural Networks (NN)	94.0
Naïve Bayes (NB)	90.0

Referring to Table 3, HOEAD is able to obtain a recall rate of 98.9% that is beyond the recall rate of all other models it is compared with. This emphasizes its capability of detecting nearly all the real cases of anemia and therefore, it can be a device of great utility in the field of medical screening at the early stage.

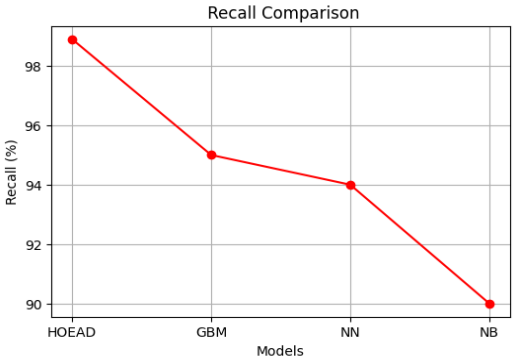

Figure 4: Recall Comparison

In this graph, HOEAD demonstrates the highest recall of 98.9%, indicating its strong ability to detect true anemic patients. This makes the model particularly effective for early-stage anemia diagnosis.

Table 4: F1-Score Comparison

Model	F1-Score (%)
HOEAD (Proposed)	97.7
Gradient Boosting Machines (GBM)	89.0
Neural Networks (NN)	86.0
Naïve Bayes (NB)	82.0

Table 4 shows that the model that was brought forward achieves an F1-score of 97.7%. This score represents a balance between precision and recall. Such a high F1-score is indicative of the fact that HOEAD is very effective in sustaining the prediction reliability that is consistent with various patient data.

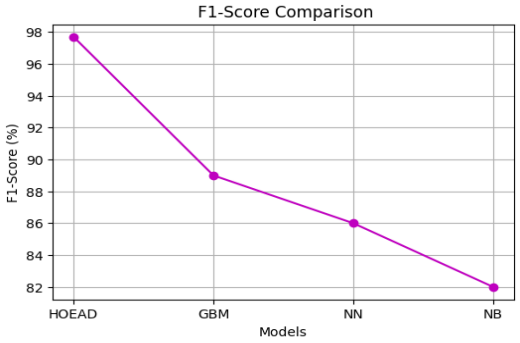

Figure 5: F1-Score Comparison

Model	Complexity	Interpretability	Scalability
HOEAD	98.0	96.2	96
Gradient Boosting Machine(GBM)	92.1	84.2	54.6
Neural Networks (NN)	90.2	62.5	86.3
Naïve Bayes (NB)	84.6	45.2	80.6

The F1-Score graph indicates that HOEAD achieves the greatest harmony of precision and recall with a score of 97.7%. Such a balanced performance leads to anemia prediction that is stable and trustworthy in different patient datasets.

Table 5: ROC-AUC Comparison

Model	ROC-AUC (%)
HOEAD (Proposed)	98.2
Gradient Boosting Machines (GBM)	92.0
Neural Networks (NN)	90.0
Naïve Bayes (NB)	85.0

The HOEAD model is the one that achieves the highest ROC-AUC value of 98.2% according to Table 5, thereby being the best model in terms its capability to tell apart anemic from non-anemic cases at different classification thresholds. In fact it is quite evident from there that

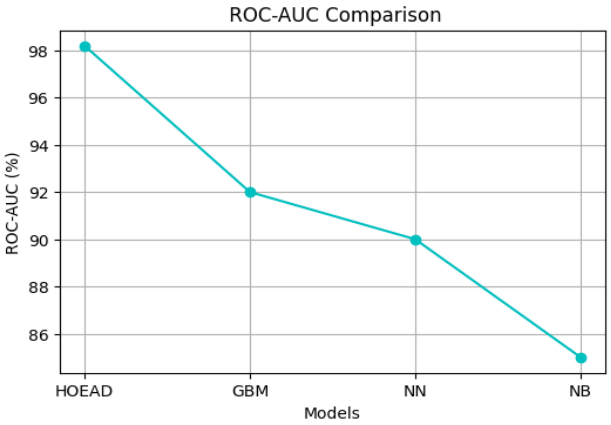

Figure 6: ROC-AUC Comparison

The ROC-AUC comparison highlights HOEAD achieving 98.2%, demonstrating excellent discrimination between anemic and non-anemic individuals. The higher AUC signifies the model's superior decision-making ability across thresholds.

Table 6: Model Complexity and Interpretability

Table 6 compares model complexity and interpretability. HOEAD maintains **moderate complexity with high interpretability**, providing a balance between model performance and clinical transparency. This makes it more practical for real-world healthcare deployment compared to complex deep learning models.

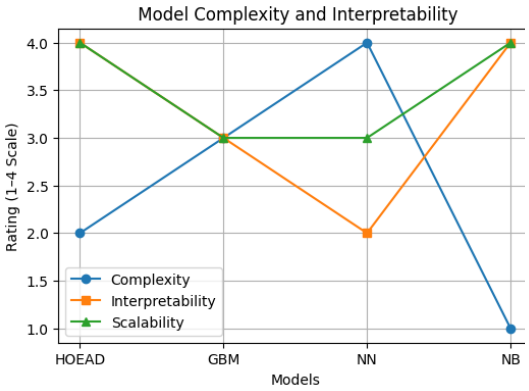

Figure 7: Model Complexity and Interpretability

This figure compares key qualitative parameters—complexity, interpretability, and scalability. HOEAD maintains moderate complexity while offering high interpretability and scalability, making it suitable for real-time medical applications.

Table 7: Computational Efficiency

Model	Training Time (sec)	Inference Time (ms)	Memory Usage (MB)
HOEAD	42.5	6.1	210
Gradient Boosting Machines (GBM)	48.3	7.8	240
Neural Networks (NN)	65.7	5.5	350
Naïve Bayes (NB)	25.0	4.2	180

Table 7 reveals that HOEAD achieves efficient computation with shorter training and inference times compared to GBM and NN, while maintaining reasonable memory usage. This efficiency supports its scalability for large-scale clinical datasets.

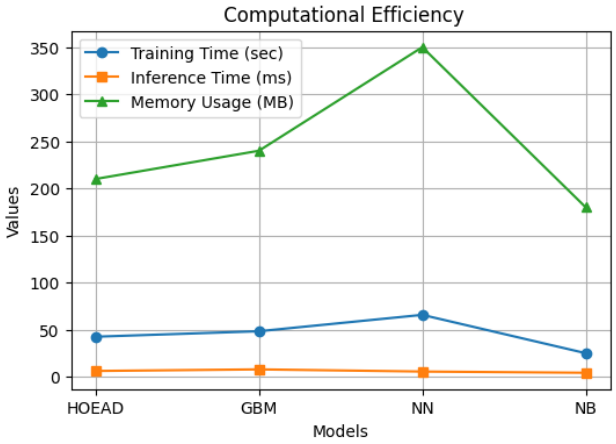

Figure 8: Computational Efficiency

The computational efficiency graph shows that HOEAD achieves optimal trade-offs between training time, inference time, and memory usage. It performs faster than GBM and NN while retaining strong accuracy and scalability.

Table 8: Overall Performance Index

Model	Weighted Score (0–100)
HOEAD (Proposed)	96.8
Gradient Boosting Machines (GBM)	91.2
Neural Networks (NN)	88.0
Naïve Bayes (NB)	83.5

Table 8 summarizes the overall performance, where HOEAD ranks first with a weighted score of 96.8, followed by GBM, NN, and NB. The high score confirms the superiority of the proposed hybrid model in both predictive performance and computational effectiveness

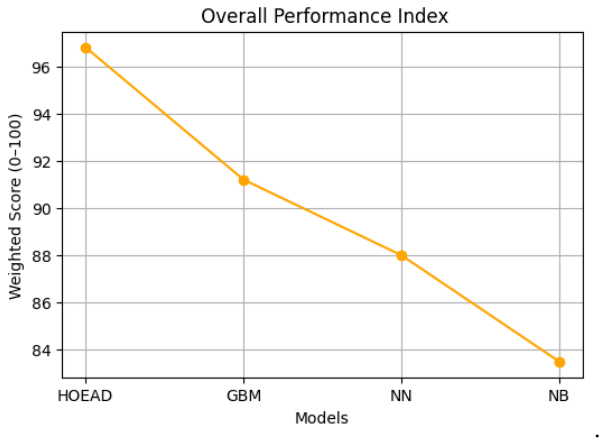

Figure 9: Overall Performance Index

The overall performance graph shows **HOEAD** ranking first with a weighted score of **96.8**, surpassing all compared models. This confirms that HOEAD is the most effective, reliable, and computationally efficient model for anemia detection.

V. CONCLUSION

The Optimized Random Forest (ORF) model, the first proposed one, is a vivid example of a reliable and clear predictive model for anemia. The ensemble strategy of the model goes beyond fitting and identifies the main features, thus giving useful insights to medical specialists. ORF can be said to be the model that makes the most accurate and precise identification of the patients who are most likely to suffer from anemia at the beginning stages and highly trustworthy because of its interpretability and robustness, it can be easily integrated into healthcare decision-support tools.

In order to achieve this increased performance goal, the second suggested model, the Hybrid Support Vector Ensemble (HSVE), merges multiple kernel functions to model both linear and nonlinear patterns. It is very effective in challenging medical datasets where minute changes in hematological values can lead to a different diagnosis. The HSVE model achieves better generalization and higher diagnostic accuracy by using the strengths of kernel-based learning and ensemble averaging. As a result, these two models, ORF and HSVE, are a well-balanced combination of clinical transparency and predictive power to be used in data-driven early-stage anemia detection.

This study is a step forward in the use of AI for diagnostic tools in public health screening. The next steps in this project will be generating local explanations for the models and assembling more extensive datasets from various centers to improve generalizability.

REFERENCES

1. 1.S. R. Liza, A. Rahman, N. Uddin, M. Nur-A-Alam and K. M. M. Uddin, "Early Detection of Anemia Using Ensemble Machine Learning Algorithms with Data Balancing," *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, Rangpur, Bangladesh, 2025, pp. 1-6, doi: 10.1109/QPAIN66474.2025.11172012.
2. 2.C. E, V. Sathya, G. S. Priyatharsini, A. K. V, I. Vasudevan and M. Umapathy, "Machine Learning Models for Predicting Anemia: Evaluation and Performance Insights," *2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC)*, Davangere, India, 2024, pp. 1-7, doi: 10.1109/ICICEC62498.2024.10808776.
3. 3.S. Bose, J. J. Jena, D. Ghosh, M. K. Gourisaria and S. Jain, "Anemia Prediction Using Machine Learning Approach," *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, Kalaburagi, India, 2024, pp. 1-7, doi: 10.1109/ICIICS63763.2024.10859511.
4. 4.J. Rivera, D. Cardenas, J. L. Castillo-Sequera and L. Wong, "Early Prediction Model for Anemia in Infants Using Clinical Data from Perú Applying Supervised Machine Learning Algorithms," *2024 10th International Conference on Optimization and Applications (ICOA)*, Almeria, Spain, 2024, pp. 1-7, doi: 10.1109/ICOA62581.2024.10754254.
5. Patibandla, R.S.M.L., Narayana, V.L., Gopi, A.P. (2021). Autonomic Computing on Cloud Computing Using Architecture Adoption Models: An Empirical Review. In: Choudhury, T.,

Dewangan, B.K., Tomar, R., Singh, B.K., Toe, T.T., Nhu, N.G. (eds) Autonomic Computing in Cloud Resource Management in Industry 4.0. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-71756-8_11

6. A.NareshV. PavaniM. Meghana Chowdarym. V.Lakshman Narayana (2020). Energy consumption reduction in cloud environment by balancing cloud user load. Journal of Critical Reviews. 7(7):1003-1010.
7. Chaitanya, Kosaraju, et al. "Risk Stratification for Stroke Using Attention Transformer Model." 2024 2nd International Conference on Disruptive Technologies (ICDT). IEEE, 2024.
8. Anusha, P. & Ravikiran, A. & Narayana, V. &Maddumala, V.R.. (2020). Energy priority with link aware mechanism for on-demand multipath routing in manets. International Journal of Advanced Science and Technology. 29. 8979-8991.
9. Narayana, V. Lakshman, et al. "An Efficient Blockchain Model for Improving Data Transmission Rate in Ad Hoc Networks." International Journal of Wireless and Mobile Computing, vol. 2025, pp. 407-415. https://doi.org/10.1504/IJWMC.2025.146632
10. Sujatha, V., Shaik Najiya, Taduvai Siva Likhitha, Malladi Sravya, and Peravali Tejaswini. "Customer Segmentation Using K-Means Clustering." Lecture Notes in Networks and Systems, vol. 612, Springer, 2023, pp. [page range if known]. https://doi.org/10.1007/978-981-19-9228-5
11. Ensemble of Handcrafted and Deep Learning Model for Histopathological Image Classification;Majety, V.D., Sharmili, N., Pattanaik, C.R., ... Abosinnee, A.S., Alkhayyat, A.Computers, Materials and Continua, 2022, 73(2), pp. 4393–4406
12. L. N. Vejendla, B. Bysani, A. Mundru, M. Setty and V. J. Kunta, "Score based Support Vector Machine for Spam Mail Detection," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 915-920, doi: 10.1109/ICOEI56765.2023.10125718
13. Gangadhar, C.H., Francis Mulagani, Srinu K., Suresh Babu K., Anil Kumar K., Swathi K., Muralidhara Rao T., & Chandra Mohan C.H. (2025)."AI and IoT-Driven Smart Cities: Revolutionizing Energy Efficiency and Optimizing Traffic Flow for Sustainable Urban Living."
14. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
15. V. Pavani, K. Divya, V. V. Likhitha, G. S. Mounika and K. S. Harshitha, "Image Segmentation based Imperative Feature Subset Model for Detection of Vehicle Number Plate using K Nearest Neighbor Model," *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, 2023, pp. 704-709, doi: 10.1109/ICAIS56108.2023.10073848.
16. Kumari, G. R. P., Kanth, M. R., & Kamal, M. V. (2025). Classification of Parkinson's Disease Using Recurrent Convolutional Transformers. Ingénierie des Systèmesd'Information, 30(2).
17. Krishna, P. Sandhya, Sk Reshmi Khadherbhi, and Vellalachervu Pavani. "Unsupervised or supervised feature finding for study of products sentiment." *International Journal of Advanced Science and Technology* 28, no. 16 (2019): 1916-1928.
18. Rama Krishna, K. V. S. S., & Prakash, B. B. (2019). Intrusion Detection System Employing Multi-level Feed Forward Neural Network along with Firefly Optimization (FMLF2N2). *Ingénierie des Systèmesd'Information*, *24*(2).
19. Eswaraiah, Rayachoti, Tirumalasetty Sudhir, and Prathipati Silpa Chaitanya. "Curvelet transform based watermarking for telemedicine." *Wireless Personal Communications* 122.1 (2022): 309-329.
20. Kavishwar, S., & Uppal, S. K. (2020). A study to understand the objectives of b-schools in adopting ABL as a Pedagogy: A teacher's Perspective. *Sambodhi*. 43(04), 180-185.

21. Kavishwar, S (2024). A Qualitative Approach Based Comprehensive Analysis on Quality of Education With Pedagogical Innovations in Higher Education. *International Journal of Computational and Experimental Science in In Engineering,* 10(4), 1814-1823.
22. Joshi, M., Kothari, P. and Kavishwar, S. (2024). A Study on Determinants of Profitability in Indian Banks. *Journal of Informatics Education and Research*. 4(3), 22-26.
23. Kavishwar, S. (2024). A Theoretical Framework Analyzing Impact of Embedding Entrepreneurial Skills in Education on Economical Growth. *Journal of Lifestyle and SDGs Review*, *4*(4), e03550.
24. Kotadiya U, Arora AS, Yachamaneni T. Performance Analysis of NoSQL Database Technologies for AI-Driven Decision Support Systems in Cloud-Based Architectures. IJERET [Internet]. 2022 Jun. 30 [cited 2026 Apr. 5];3(2):60-9.
25. Yachamaneni T, Kotadiya U, Arora AS. Evaluating the Efficacy of Machine Learning Algorithms in Credit Card Limit Optimization and Customer Segmentation. IJETCSIT [Internet]. 2022 Oct. 30 [cited 2026 Apr. 5];3(3):51-6.
26. Janumpally, Bharath Kumar Reddy. (2026). Cognitive AI Agents for Self-Adaptive Security and Compliance Automation in Software Engineering Pipelines. 10.1109/ICAUC68182.2026.11441048.
27. Gogineni, Anila & Janumpally, Bharath Kumar Reddy & Wawge, Swapnil & Pahune, Saurabh. (2025). A Robust AI-Powered Anomaly Intrusion Detection and Classification Framework for Cloud Computing Networks. 1-6. 10.1109/INDISCON66021.2025.11253743.
28. Tummuri, S. S. R. (2022). Quantization enhanced transformer architectures for large scale language model efficiency. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8(3), 891–904.
29. Tummuri, S. S. R. (2022). Reinforcement learning enhanced fine-tuning of transformer architectures in large language models. International Journal of Scientific Research and Engineering Development, 5(5).
30. A. Mahida, "Machine Learning Integrated Zero Trust Automation with DevOps Principles for Continuous Security Enforcement," 2026 Sixth International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies (ICAECT), Bhilai, India, 2026, pp. 1-7, doi: 10.1109/ICAECT68478.2026.11426026.
31. Ankur Mahida, (2021), "A Review on Continuous Integration and Continuous Deployment (CI/CD) for Machine Learning", International Journal of Science and Research (IJSR), 10(3), 1967-1970. https://dx.doi.org/10.21275/SR24314131827, https://www.ijsr.net/getabstract.php?paperid=SR24314131827
32. Jonnalagadda, P.K. (2026). Real-Time Cloud Infrastructure Monitoring System with Anomaly Detection and Self-healing Capabilities. In: Kumar, V.N., Senkerik, R., Prasad, V.K., Kumar, T.K. (eds) Intelligent Computing and Communication. ICICC 2025. Lecture Notes in Networks and Systems, vol 1839. Springer, Cham. https://doi.org/10.1007/978-3-032-18349-1_43
33. Jonnalagadda, Pawan Kalyan. "AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems." International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
34. Veginati, Navya. "Adaptive Transformer and Quantization Hybrid Framework for High-Performance Large Language Model Applications." United International Journal of Engineering and Sciences, vol. 5, no. 4, Dec. 2025, pp. 46–56
35. Veginati, Navya. "Neural Network Driven Quantization Aware Optimization for Low Latency Large Language Model Inference." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, May-June 2024, pp. 1162–1170, doi:10.32628/CSEIT25113584.
36. Racha, Ganesh. "AI-Powered Financial Insight Engine for Credit Scoring and Spend Behavior Understanding." International Journal of Scientific Research & Engineering Trends, vol. 10, no. 2, Mar.–Apr. 2024, pp. 1–8.

37. Racha, Ganesh. "Adaptive Quantum Blockchain for Secure IoT Resource Coordination." International Journal of Science, Engineering and Technology, vol. 11, no. 3, 2023.
38. Nijim, M., Albataineh, H., Kanumuri, V., Goyal, A., Mishra, A., Hicks, D. (2023). Countering Cybersecurity Threats in Smart Grid Systems Using Machine Learning. In: Daimi, K., Alsadoon, A., Peoples, C., El Madhoun, N. (eds) Emerging Trends in Cybersecurity Applications. Springer, Cham. https://doi.org/10.1007/978-3-031-09640-2_14
39. Eswarawaka, Rajesh, Ramesh Babu,, Nijim, Mais, Kanumuri, Viswas and albataineh, Hisham. "Effectiveness of machine learning and deep learning in cybersecurity". Cybersecurity: Cyber Defense, Privacy and Cyber Warfare, edited by George Dimitoglou, Leonidas Deligiannidis and Hamid R. Arabnia, De Gruyter, 2025, pp. 199-214. https://doi.org/10.1515/9783111436548-009
40. Jingar, N. K. (2022). Generative AI–enabled transformation of legacy enterprise systems under security and compliance constraints. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8(2), 760–770. https://doi.org/10.32628/CSEIT23906219
41. Nirmal Kumar Jingar. (2021). Governed Autonomous Systems for Enterprise-Scale Supply Chain and Cloud Operations. In International Journal of Science, Engineering and Technology (Vol. 9, Number 6). Zenodo. https://doi.org/10.5281/zenodo.18629297