Developing A Hybrid Machine Learning Algorithm for Anemia Diagnosis

Safa S. Abdul-Jabbar [1] *, Alaa k. Farhan[2], Abdul Hafeez Kandhro[3]
[1] Computer Science Department/ College of Science for Women University of Baghdad, Baghdad, Iraq.
[2] Computer Science Department, University of Technology, Baghdad, Iraq.
[3]Institute of Medical Technology, Jinnah Sindh Medical University, Karachi, Pakistan

Emails: Safa.s@csw.uobaghdad.edu.iq, alaa.k.farhan@uotechnology.edu.iq, abdulhafeez.kandhro@jsmu.edu.pk

Abstract

The utilization of artificial intelligence (AI) algorithms has significantly transformed the field of blood disease diagnosis, enabling enhanced capabilities in prediction, categorization, and optimization. However, there is still a lack of research exploring the advancement of hybrid machine learning models that combine qualitative and quantitative datasets to address issues associated with blood diseases. To tackle this gap, we evaluate algorithmic combinations using datasets that include key characteristics from complete blood count (CBC) examinations. This manuscript presents an evaluation of prominent deep learning models, such as CNN, RNN, and RCNN, as part of our methodology. The assessment identified XGBoost as the optimal machine learning algorithm, and RCNN as the best deep learning model. Consequently, we propose a hybrid model named 'RCNNX,' which integrates Robust Scaler, SelectKBest feature selection, RCNN, and the XGBoost algorithm. The hybrid model, 'RCNNX,' achieves exceptional testing accuracy levels of 100% and 95.12% on the Anemia Diagnosis Dataset and a second dataset, respectively. Additionally, it demonstrates recall rates of 100% and 94.64% for the same datasets. These findings highlight the superiority of the proposed model, as it effectively utilizes feature selection to reduce the number of input variables, minimizing the risk of overfitting. Moreover, XGBoost enhances the predictive accuracy of RCNN.

Keywords: Complete Blood Cell Count; CBC Test Parameters; Machine Learning Algorithms; Anemia Classifications; Blood Diseases Prediction.

1 Introduction

Anemia is a widespread condition characterized by a lower-than-normal hemoglobin concentration in the blood. It can significantly impact an individual's overall health and well-being, causing fatigue, weakness, dizziness, shortness of breath, and an increased risk of infections [1]. Anemia is a major public health concern, affecting a large portion of the global population. It can result from nutrient deficiencies, genetic disorders, chronic diseases, and certain medications. Thus, early and accurate diagnosis and treatment are essential to preventing complications and improving the quality of life for those affected [2].

In medicine, AI is focused on designing highly accurate algorithms and diagnostic methods. Diagnosing a medical condition involves identifying the diseases or conditions responsible for a patient's signs and symptoms [3]. Despite advancements, there remain unmet needs worldwide in the accuracy of early disease detection, particularly when signals are complex and patients exhibit diverse underlying symptoms—posing a significant challenge for healthcare professionals developing diagnostic tools [4].

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

To create a highly efficient and accurate AI model, data preprocessing is essential, involving multiple stages that enhance input data and contribute significantly to building precise models. Feature scaling and selection are critical components that play a crucial role in improving prediction efficiency and should not be overlooked [5]. This framework discusses how these elements synergistically enhance the diagnostic process, emphasizing their combined effects rather than their technical aspects alone [6,7]. The integration of RCNN aims to maximize the strengths of both RNN and CNN, alongside XGBoost and other techniques. Furthermore, XGBoost's ability to handle diverse datasets, perform well across various challenges, and produce accurate predictions with low computational cost is a key advantage in this model [8,9].

2 Literature Reviews

Anemia is a widespread condition characterized by a lower-than-normal hemoglobin concentration in the blood. It can significantly impact an individual's overall health and well-being, leading to fatigue, weakness, dizziness, shortness of breath, and an increased risk of infections. As a major public health concern affecting a large portion of the global population, anemia can result from nutrient deficiencies, genetic disorders, chronic diseases, or certain medications. Therefore, early and accurate diagnosis and treatment are essential to preventing complications and improving the quality of life for those affected [4].

- In medicine, AI focuses on designing highly accurate algorithms and diagnostic methods. Diagnosing a medical condition involves identifying the diseases or conditions responsible for a patient's signs and symptoms. Despite advancements, unmet needs remain worldwide in the accuracy of early disease detection, especially when signals are complex and patients exhibit diverse underlying symptoms—posing significant challenges for healthcare professionals developing diagnostic tools [10].
- To create a highly efficient and accurate AI model, data preprocessing is crucial. This involves multiple stages that enhance input data and contribute significantly to building precise models. Feature scaling and selection are critical components that improve prediction efficiency and should not be overlooked. This framework discusses how these elements synergistically enhance the diagnostic process, focusing on their combined effects rather than just their technical aspects. The integration of RCNN aims to maximize the strengths of both RNN and CNN, alongside XGBoost and other techniques. Additionally, XGBoost's ability to handle diverse datasets, perform well across various challenges, and produce accurate predictions with low computational cost is a key advantage in this model.

The most significant findings from the aforementioned studies are presented in Table 1.

Table 1: Hematological Diagnostics Literature Review Summarization

Ref.	Year	Main Methodology	Problem Statements	Main Contribution	Accuracy	Limitations
Kandhro, 2017	2017	Random Forests and Decision Tree	Distinguishing between TTs and IDA	• Improve a cutoff value utilized by the RF and decision-tree approach. • Develop a unique formula Kandhro1 & Kandhro2 for differentiating between TTs and IDA.	100%	Dataset size: 551 TTs, 441 IDA
Jaiswal, 2018	2019	Naive Bayes, Random Forest, and Decision Tree	Determine the appropriate algorithm for the CBC dataset	• Provide automated tools to diagnose anemia. • Approve that Naive Bayes achieved 96.09% accuracy.	96.09%	Dataset size: 200 records
Alsheref, 2019	2019	LogitBoost, Random Forests, Decision Tree, Regression Analysis, KNN, Bayesian Network, Multilayer Perceptron, Naive Bayes, SVM	Determine the appropriate algorithm for the CBC dataset	• Provide information on the best algorithm for blood data prediction. • Aid doctors in diagnoses based on CBC results.	98.16%	Dataset size: 668 records
Haider, 2022	2022	Artificial Neural Network	Develop an accurate model to diagnose hematological emergencies	• Provide accuracy for training/testing datasets. • ANN utilizes hidden trends in disease signatures.	83.1% (training), 89.47% (testing)	Clinical usefulness limited
Vohra, 2022	2022	Decision Tree, Logistic Regression, Multilayer Perceptron, Naive Bayes, Random Forest, SVM	Detection of Anemia at various stages	• Compare performance of ML algorithms. • Use three feature selection methods (filter, wrapper, embedded).	94.44%	Dataset size: 400 records

3 Theoretical Background

In this section, the foundational knowledge and concepts essential for designing an AI model for disease detection are presented. It also establishes the necessary framework and understanding for the subsequent research steps, which include:

https://doi.org/10.54216/JAIM.090103
Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

3.1 Data Preprocessing

In this step, feature selection and scaling were applied to preprocess the data before it was fed into the proposed model.

3.1.1 Feature Scaling

Feature scaling is a method of adjusting data values to fit within a defined range, allowing the model to perform more effectively and improve its overall performance. Depending on the data and the specific problem, various techniques can be employed for scaling input features, including [11, 12]:

- **Min-Max Scaling:** This technique scales features to fit within a defined range, typically 0 to 1.
- **Robust Scaling:** Similar to Min-Max scaling, but it uses the interquartile range instead of a specific range, improving resistance against outliers.
- **Normalization:** This technique scales features so that their unit norms are equal to 1. It is particularly useful when the magnitude of features is irrelevant, and only their direction matters, as in certain clustering algorithms.

3.1.2 Feature Selection

Feature selection involves identifying and choosing the most significant features that contribute effectively to prediction tasks. It decreases the number of input variables to make the model more efficient. Commonly used feature selection techniques include [13, 14]:

- **Filter Methods:** These select features based on statistical methods such as the chi-squared test, correlation coefficient, mutual information, SelectKBest, and variance threshold.
- **Wrapper Methods:** These select features based on a model's performance using approaches like recursive feature elimination (RFE), forward selection, and backward elimination.
- **Embedded Methods:** These integrate feature selection into the model-building process, such as Lasso regression, decision trees, and random forests.
- **Dimensionality Reduction:** This involves mapping high-dimensional data into a lower-dimensional space while retaining as much information as possible, using techniques like principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE).

3.2 ML and DL Models

AI has the potential to revolutionize disease diagnosis by offering more accurate and efficient methods [15]. However, several obstacles must be overcome to achieve this goal. Several pivotal models can be employed to address these challenges:

3.2.1 Machine Learning Models

Machine Learning (ML) analyzes data samples and draws conclusions using mathematical and statistical methods. Popular ML models for diagnosing anemia include Random Forest (RF) and Multilayer Perceptron (MLP), known for effectively analyzing large datasets, identifying patterns, and generating accurate predictions. Studies have demonstrated their robustness and reliability for anemia diagnosis [16, 17]. Additionally, ensemble methods like XGBoost have achieved high accuracy in disease prediction [18].

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

3.2.2 Deep Learning Models

Deep Learning (DL), a subfield of Machine Learning, simulates the way the human brain processes information. Originating from Artificial Neural Networks (ANN), DL has gained prominence in healthcare, visual recognition, text analytics, and other fields [19–21]. Common representations of real-world data for DL modeling include:

- **Sequential Data:** Data where the order is critical.
- **Two-dimensional Data (Images):** A grid of numbers, symbols, or phrases in rows and columns.
- **Tabular Data:** Primarily consists of rows and columns.

Popular deep neural networks include Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) [22]. Additionally, the Recurrent Convolutional Neural Network (RCNN) combines convolutional and recurrent layers by first using a CNN to extract features and then a recurrent layer to describe temporal relationships [23].

When working with nonlinear transition functions in Markov models, modifications like Long Short-Term Memory (LSTM) are used to capture temporal dependencies effectively. LSTMs outperform traditional RNNs by incorporating three gates (forget, input, and output) within each neuron [24]. Optimization algorithms further enhance training efficiency, with popular choices including Stochastic Gradient Descent (SGD), Adam, AdaDelta, and others [25, 26].

Table 2: Comparison Summary of CNN, RNN, and RCNN Models

Property	**CNN**	**RNN**	**RCNN**
Type of Data	Image Data	Sequential Data	Image and Sequential Data
Parameter Sharing	Yes	Yes	Yes
Fixed Length Input	Yes	No	No
Recurrent Connections	No	Yes	Yes
Spatial Relationship	Yes	No	Yes
Temporal Relationship	No	Yes	Yes
Main Advantages	Superior performance in image recognition	Predicts time series	Combines RNN and CNN strengths
Disadvantages	Requires a lot of training data	Gradient vanishing/exploding	—
Application	NLP, facial recognition	Text-to-speech	Image/video object detection

4 Proposed Model Design

Using machine learning algorithms that learn to label instances as belonging to a particular class is crucial for the classification step in all diagnostic and prediction systems.

In this paper, 11 different algorithms were applied as a first step: LogitBoost (Logit), Random Forest (RF), Decision Tree (DT), XGBoost (XGB), Gradient Boosting (GB), Multilayer Perceptron (MLP), AdaBoost (Ada), Logistic Regression (LR), Support Vector Machine (SVM), CatBoost (Cat), and K-Nearest Neighbor (KNN). These algorithms are widely used due to their effectiveness in processing large datasets, detecting patterns and correlations, and producing accurate predictions. Additionally, the robustness and reliability of these algorithms for anemia diagnosis have been extensively demonstrated in numerous studies, making them popular among healthcare practitioners and hematology researchers.

4.1 Algorithm Selection and Refinement

Several experiments were conducted with different parameter values, feature scales, and selection criteria to refine the algorithms and improve their precision. Based on the results, Logit, RF, XGBoost, LR, and MLP were identified as the most suitable algorithms for achieving optimal performance. Advanced techniques such as feature scaling and selection were employed, including:

https://doi.org/10.54216/JAIM.090103
Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

- **Feature Scaling:** Robust Scaler for feature standardization.
- **Feature Selection:** Select K-Best method using the F-score.

Hyperparameter tuning was performed to further enhance algorithm performance.

4.2 Deep Learning Model Comparison

In addition to machine learning models, we evaluated several widely adopted deep learning models, such as CNN, RNN, and RCNN. These algorithms were tested on three international datasets, as detailed in Section 5.1. After rigorous evaluation, RCNN emerged as the prime candidate for the new hybrid model due to its superior testing outcomes.

4.3 Proposed Hybrid Model: RCNNX

The proposed hybrid model combines RCNN, feature scaling, feature selection, and the XGBoost algorithm. The integration of XGBoost enhances the accuracy of multi-class classification tasks performed on the outputs of the RCNN model. The RCNN model generates predicted probabilities for each class, and XGBoost is trained with these probabilities to refine predictions. This approach leverages the strengths of RCNN for feature extraction and XGBoost for classification, resulting in more precise predictions.

4.4 Algorithm Steps: RCNNX

The detailed steps of the proposed hybrid algorithm, RCNNX, are outlined in Algorithm 4.4.

[H] **Input:** Input dataset (Excel file)
Output: Loss and accuracy plots for training and testing sets, saved models.

1. Load and split the data into feature matrix X and target vector y.
2. Perform data preprocessing:
   - Encode labels.
   - Handle missing values.
3. Split X into training, testing, and validation sets.
4. Scale training and test sets using the Robust Scaler.
5. Select top K features using Select K-Best with the F-score function.
6. Remove non-selected features from training and testing data.
7. Reshape input data for the RCNN model.
8. Build the RCNN model:
   - Add a bidirectional LSTM layer.
   - Add a convolutional layer.
   - Add a pooling layer.
   - Add a flatten layer.
   - Add an output layer with Softmax activation.

https://doi.org/10.54216/JAIM.090103
Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

9. Compile the RCNN model using the Adam optimizer, categorical cross-entropy loss function, and accuracy metrics.
10. Define callbacks for:
    - Early stopping.
    - Model checkpointing.
    - Reducing learning rate.
11. Train and validate the RCNN model.
12. Extract predicted probabilities from the RCNN model.
13. Reshape the predicted probabilities into a 2D matrix.
14. Define a custom objective function for XGBoost using the gradient and Hessian.
15. Train the XGBoost model on the predicted probabilities.
16. Test and evaluate the XGBoost model's performance.
17. Save the trained RCNN and XGBoost models.
18. Plot accuracy and loss results.

4.5 Proposed Model Architecture

The RCNNX model architecture incorporates a bidirectional Long Short-Term Memory (LSTM) layer, convolutional layers, pooling layers, flattening layers, and output layers. This design in Figure 1 effectively integrates the strengths of CNN and RNN technologies. Additionally, XGBoost is used for its ability to handle high-dimensional datasets efficiently, perform well across various challenges, and produce accurate predictions at low computational cost.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

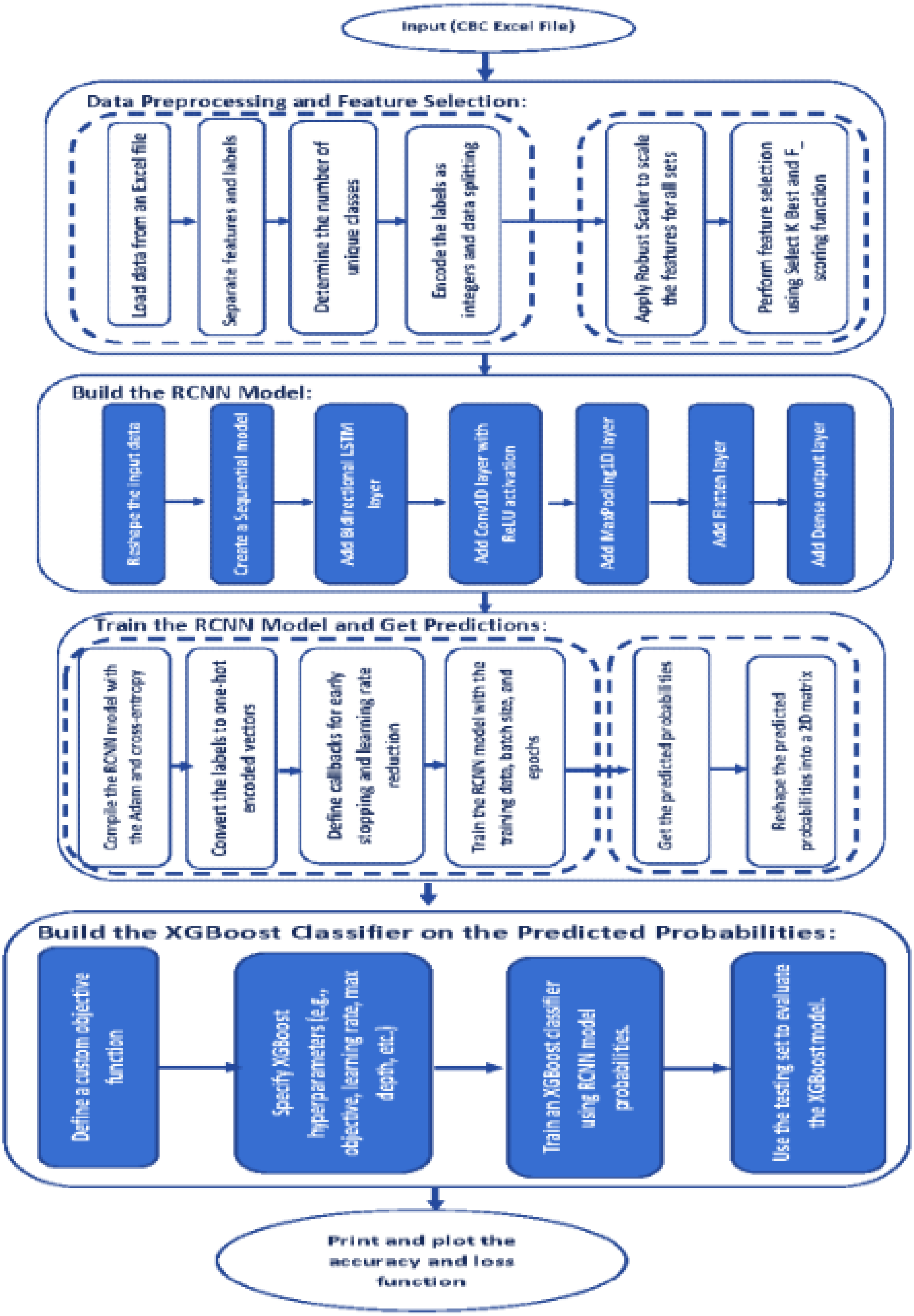

Figure 1: Diagram of the proposed learning model (RCNNX).

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

5 Experimental Results and Discussion

The strengths of both ML and DL techniques were combined to create a hybrid model that leverages their respective advantages. The following sections present the outcomes of this hybridization process and the findings from the comparative study of each approach.

5.1 Dataset Description

The CBC datasets used in this study contain several basic transactions common to hematology research. Two datasets were employed to train and test the proposed model, as described in Table 3. Notably, 'Dataset2' is a private dataset focusing on three years (2012–2015) of encoded and unlinked clinical laboratory data obtained from the Healthcare Molecular and Diagnostic Laboratory in Hyderabad, Pakistan.

Table 3: Descriptions of International Tested Datasets

Author	Dataset Name	Year	Disease Type	Country	No. of Samples	No. of Features	Dataset Type
Shahane, 2020	Anemia Diagnosis Dataset	2021	Anemia	India	1442	5	Open Access
Kandhro, 2017	Dataset2	2017	Anemia	Pakistan	410	21	Closed Access

5.2 Machine Learning Comparative Study Results

To rigorously test commonly used ML algorithms and identify the most suitable ones for hyperparameter tuning, two international datasets were utilized. The Anemia Diagnosis Dataset consists of two classes (anemia and non-anemia) with 1,442 records. Figure 2 shows the accuracy metrics for the training phase, and Table 4 summarizes the performance of the algorithms on 25% of the dataset for testing.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

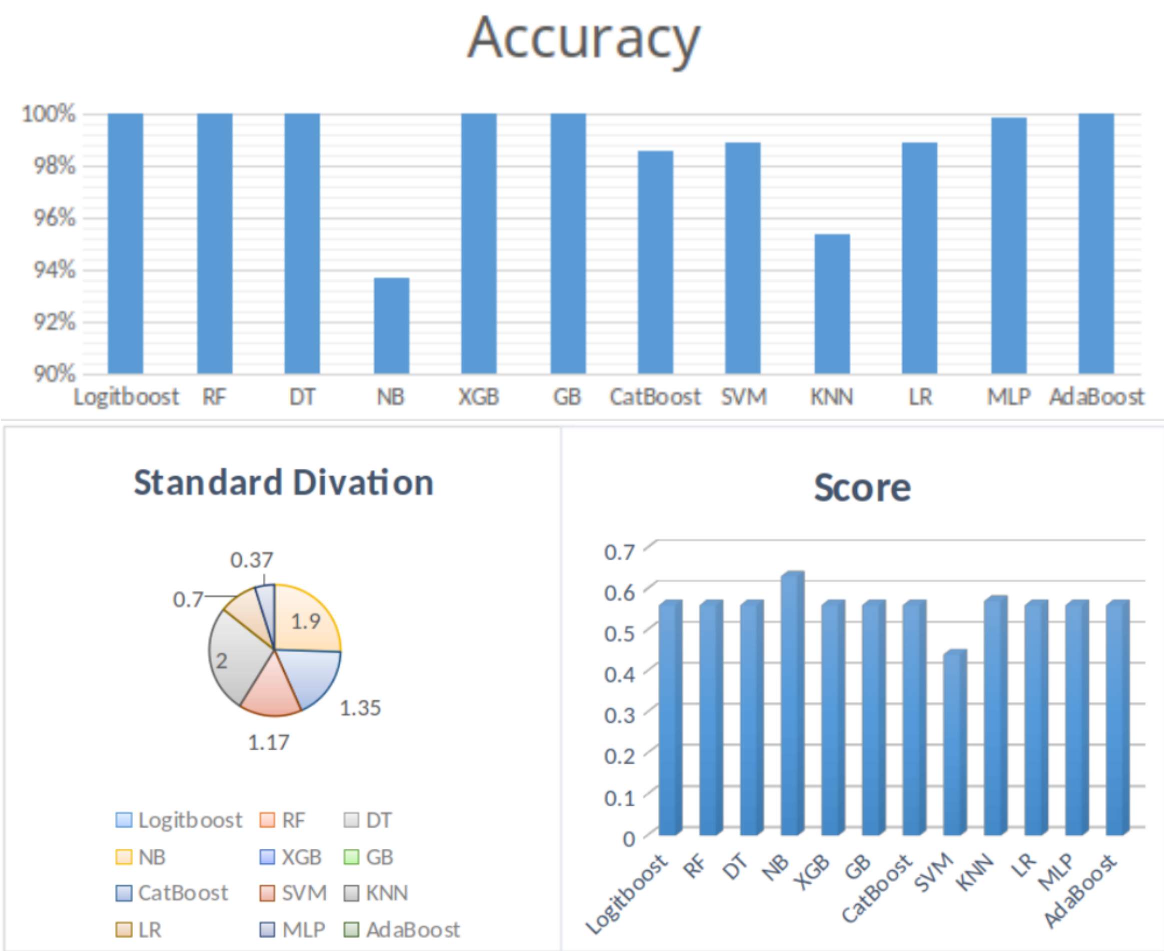

Figure 2: Accuracy metrics for the training phase using the Anemia Diagnosis Dataset (10-fold cross-validation).

Table 4: Accuracy Metrics for 25% of the Anemia Diagnosis Dataset (Testing)

Algorithm	Accuracy	Recall	Score	TNP	TRP	NPV	PPV
LogitBoost	1	1	66.66	1	1	1	1
RF	1	1	66.66	1	1	1	1
DT	1	1	66.66	1	1	1	1
NB	93.82	96.95	65.97	89.93	96.95	95.97	92.27
XGB	1	1	66.66	1	1	1	1
GB	1	1	66.66	1	1	1	1
CatBoost	99.43	99.04	66.45	1	99.04	98.65	1
SVM	99.71	1	66.66	99.33	1	1	99.51
KNN	97.19	99.00	66.44	94.83	99.00	98.65	96.13
LR	99.43	1	66.66	98.67	1	1	99.03
MLP	1	1	66.66	1	1	1	1
AdaBoost	1	1	66.66	1	1	1	1

Next, Dataset2, which contains the largest number of features (21) and 410 records, was analyzed. Accuracy metrics for the training phase are shown in Figure 3, while Table 5 provides the results for testing.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

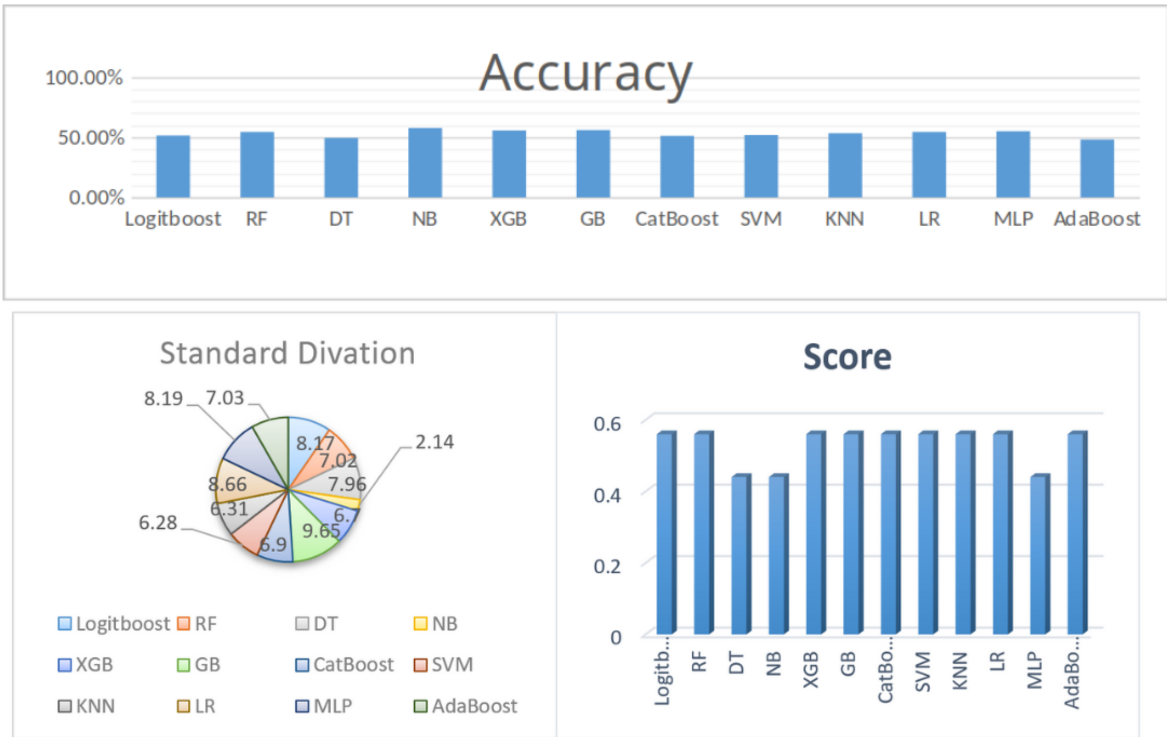

Figure 3: Accuracy metrics for the training phase using Dataset2 (10-fold cross-validation).

Table 5: Accuracy Metrics for 25% of Dataset2 (Testing)

Algorithm	Accuracy	Recall	Score	TNP	TRP	NPV	PPV
LogitBoost	52.04	52.71	51.32	51.03	52.71	41.67	61.91
RF	47.16	48.89	49.44	42.43	48.89	23.34	69.85
DT	49.51	52.45	51.20	45.23	52.45	39.58	58.18
XGB	54.36	58.57	53.94	45.45	58.57	34.09	69.49
NB	53.39	53.46	51.67	50	53.46	20.83	69.18
GB	51.45	53.84	51.85	47.36	53.84	37.50	63.63
CatBoost	45.63	49.33	49.66	35.71	49.33	20.83	67.27
SVM	48.54	51.21	50.60	38.09	51.21	16.66	76.36
KNN	46.60	50.00	50.00	38.70	50.00	25.00	65.45
LR	52.42	53.57	51.72	47.36	53.57	18.75	81.81
MLP	52.33	54.16	52.00	48.57	54.16	34.00	68.42
AdaBoost	52.42	54.28	52.05	48.48	54.28	33.33	69.09

5.3 Deep Learning Comparative Study Results

To evaluate the performance of deep learning (DL) models, the Sequential model was trained for 100 epochs. Data preprocessing included:

- Standard Scaler for feature scaling.
- Select K-Best for feature selection.
- Adam optimizer with a learning rate of 0.001.

The Long Short-Term Memory (LSTM) architecture was employed for both Recurrent Neural Network (RNN) and Recurrent Convolutional Neural Network (RCNN) models, each configured with 64 units. The results are summarized in Table 6.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

Table 6: Results of All Datasets Tests Using Well-Known Models

Model	Anemia Diagnosis Dataset (%)	Dataset2 (%)
CNN	99.32	53.48
RNN	100.00	55.81
RCNN	100.00	58.13

The results indicate that the RCNN model outperformed other models for both datasets, achieving the highest accuracy.

5.4 Proposed Model Results

The proposed hybrid model, RCNNX, was designed to achieve both high accuracy and minimal execution time. Feature scaling and feature selection were employed to enhance the performance of the RCNN model, utilizing techniques such as:

- Min-Max scaling, Robust scaling, and Normalization.
- Feature selection using Filter, Wrapper, Dimensionality Reduction, and Embedded methods.

Table 7 presents the results of the proposed model for both datasets, achieved using a 10% validation set, 20% testing set, and 70% training set.

Table 7: Results of All Datasets Using the Proposed Model (RCNNX)

Dataset	Accuracy (%)	Recall (%)
Anemia Diagnosis Dataset	100.00	100.00
Dataset2	95.12	94.64

5.5 Comparison with Prior Studies

To validate the effectiveness of the proposed model, a comparative analysis with previous studies was conducted. Table 8 highlights that the proposed model outperforms earlier works, demonstrating higher accuracy for larger datasets.

Table 8: Comparison of Proposed Model with Prior Studies

Reference	Year	No. of Records	Main Techniques	Test Accuracy (%)
Jaiswal, 2018	2019	200	Naive Bayes, RF, DT	96.09
Alsheref, 2019	2022	1577	ANN	89.47
Vohra, 2022	2022	400	Cross-validation with ML models	94.44
Proposed Model	2023	1442 410	Robust Scaler, Select K-Best, RCNN, XGBoost	100.00 95.12

5.6 Visualization of Model Performance

The training and validation phases for all datasets were monitored using accuracy and loss plots. Figure 4 illustrates these metrics for the Anemia Diagnosis Dataset and Dataset2.

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

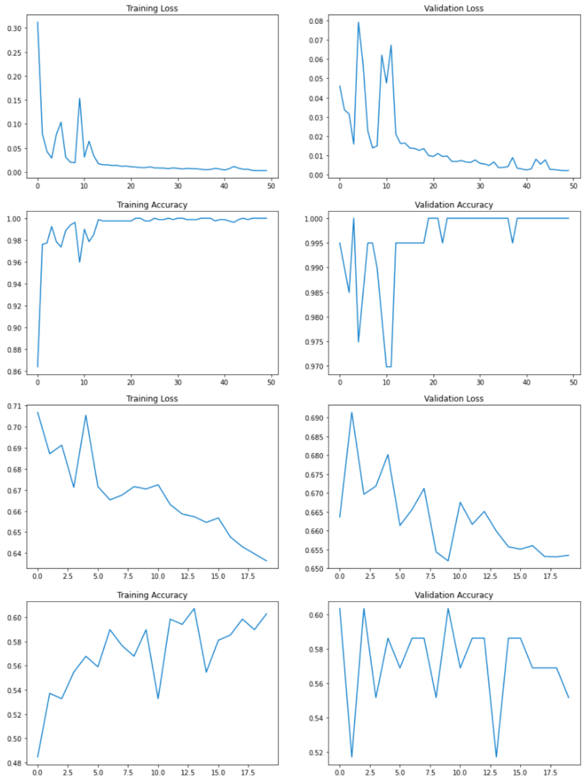

Figure 4: Training and Validation Loss and Accuracy for All Datasets.

6 Conclusions

Healthcare systems have significantly advanced, transforming medical care and becoming a crucial facet of data science research. The exploration of pathological report analysis, particularly within the realms of deep learning (DL) and machine learning (ML), is a rapidly growing area driven by the exponential growth of computational power and data availability. This paper focused on utilizing DL and ML for Comprehensive Blood Count (CBC) test data analysis, aiming to dissect and contextualize information within test results.

A hybrid model was developed to assist medical practitioners in diagnosing anemia. The XGBoost classifier, which achieved the highest performance across all evaluated datasets, was integrated with the RCNN model to create a new hybrid model that balances accuracy and processing time. The Random Forest algorithm was employed to determine the importance of each feature in the dataset, helping establish thresholds for feature selection. XGBoost was then applied as an ensemble algorithm to select the best set of features. This approach was chosen for its flexibility, scalability, and minimal requirement for feature normalization, which reduces preprocessing time.

The proposed hybrid AI model, named RCNNX, combines the following techniques:

- **Robust Scaler and Select K-Best:** For feature scaling and selection, ensuring high model efficiency and reducing dimensionality.
- **RCNN:** For extracting meaningful temporal and spatial patterns from input data.
- **XGBoost:** For improving classification accuracy and handling high-dimensional datasets.

The RCNNX model achieved remarkable accuracies of 100% and 95.12% on the Anemia Diagnosis Dataset and Dataset2, respectively, as highlighted in Table 7. These results demonstrate that the integration of XG-Boost with the RCNN model effectively manages expansive, high-dimensional datasets, optimizes decision boundaries, and delivers precise predictions.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

6.1 Future Work

While the proposed RCNNX model showcases excellent performance, there are areas for potential improvement and exploration:

- **Dataset Diversity:** Incorporating larger and more diverse datasets to generalize the model across different populations.
- **Feature Engineering:** Exploring advanced feature engineering techniques to further improve classification accuracy.
- **Real-Time Applications:** Investigating the feasibility of deploying the RCNNX model for real-time anemia diagnostics in clinical settings.
- **Explainability:** Enhancing the model's interpretability to provide medical professionals with clearer insights into predictions.

This research highlights the potential of hybrid AI models to revolutionize disease diagnosis, particularly in leveraging the strengths of ML and DL techniques. The RCNNX model serves as a robust tool for analyzing CBC test data, paving the way for improved diagnostic accuracy and patient care.

References

[1] I.H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):1–20, 2021.

[2] L.R. Al-Khazraji, A.R. Abbas, and A.S. Jamil. Generating various deep dream images through maximizing the loss function of particular layers using inception-v3 and inception-resnet-v2 models. *Iraqi Journal of Science*, pages 3468–3483, 2024.

[3] S. Shahane. Anemia diagnosis dataset. Kaggle, 2020.

[4] L.R. Al-Khazraji, A.R. Abbas, A.S. Jamil, and A.J. Hussain. A hybrid artistic model using deep-dream model and multiple convolutional neural networks architectures. *IEEE Access*, 2023.

[5] E. Stevens, L. Antiga, and T. Viehmann. *Deep learning with PyTorch*. 2020.

[6] R.-Y. Sun. Optimization for deep learning: An overview. *Journal of Operations Research Society of China*, 8(2):249–294, 2020.

[7] S. Suner, J. Rayner, I.U. Ozturan, G. Hogan, C.P. Meehan, A.B. Chambers, J. Baird, and G.D. Jay. Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera. *PLoS One*, 16:e0253495, 2021.

[8] R. Vohra, A. Hussain, A. Dudyala, J. Pahareeya, and W. Khan. Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting. *PLoS One*, 17(7):1–18, 2022.

[9] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu. Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2016.

[10] L.R. Ali and A.S. Jamil. Effect of changing targeted layers of the deep dream technique using vgg-16 model. *International Journal of Online and Biomedical Engineering*, 19(3), 2023.

[11] M. Ahsan, M.A.P. Mahmud, P.K. Saha, K.D. Gupta, and Z. Siddique. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9:5–9, 2021.

[12] M.M. Ahsan, S.A. Luna, and Z. Siddique. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare*, 10(541):1–30, 2022.

https://doi.org/10.54216/JAIM.090103

Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025

[13] F.K. Alsheref. Blood diseases detection using classical machine learning algorithms. *Journal of Biomedical Science and Engineering*, 10(7):77–81, 2019.

[14] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz. A comparative analysis of xgboost. *arXiv preprint*, 2019.

[15] N. Çolakoğlu and B. Akkaya. *Recent advances in data science and business analytics*. Mimar Sinan Fine Arts University Publications, 2019.

[16] R. Cui, H. Liu, and C. Zhang. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7361–7369, 2017.

[17] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(7):2121–2159, 2011.

[18] R.Z. Haider, I.U. Ujjan, N.A. Khan, and E. Urrechaga. Beyond the in-practice cbc: The research cbc parameters-driven machine learning predictive modeling for early differentiation among leukemias. *Diagnostics*, 12(1):138, 2022.

[19] A.A. Ibrahim, R.L. Ridwan, M.M. Muhammed, R.O. Abdulaziz, and G.A. Saheed. Comparison of the catboost classifier with other machine learning methods. *Journal of Intelligent & Fuzzy Systems*, 11:1–7, 2020.

[20] M. Jaiswal, A. Srivastava, and T.J. Siddiqui. Machine learning algorithms for anemia disease prediction: Select proceedings of ic3e 2018. In *IC3E 2018 Proceedings*. Springer, 2019.

[21] A. Jović, K. Brkić, and N. Bogunović. A review of feature selection methods with applications. In *38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1200–1205, 2015.

[22] A.H. Kandhro, W.S.V. Nuchnoi, and P. Pornlada. New bioinformatics-based discrimination formulas for differentiation of thalassemia traits from iron deficiency anemia. *Lab Med*, 48(3):230–237, 2017.

[23] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint*, 2015.

[24] J. Li, H. Zhang, J. Zhao, X. Guo, W. Rihan, and G. Deng. Embedded feature selection and machine learning methods for flash flood susceptibility-mapping in the mainstream. *Remote Sensing*, 14(21):5523, 2022.

[25] K. Maharana, S. Mondal, and B. Nemade. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, pages 91–99, 2022.

[26] P.O. Pinheiro and R.C. Com. Recurrent convolutional neural networks for scene labeling. In *International Conference on Machine Learning (PMLR)*, pages 82–90, 2014.

https://doi.org/10.54216/JAIM.090103
Received: November 02, 2024 Revised: December 26, 2024 Accepted: January 20, 2025