A Deep Learning Approach for Prediction of Fetal Hemoglobin Reactivation and Stratify Disease Severity in β Thalassemia

Md. Moshfiqur Rahaman[1*]

[1*]Dhaka, Bangladesh.

Corresponding author(s). E-mail(s): mushfiqshahed@gmail.com;

Abstract

Reduced hemoglobin production is the hallmark of $\boldsymbol{\beta}$ thalassemia, a genetic blood condition that causes anemia of variable severity. The treatment of fetal hemoglobin (HbF) reactivation is essential for reducing the symptoms of the disease. This study suggests a novel Multi-Task Attention Transformer (MT-AT) deep learning model that uses extensive clinical and genetic patient data to stratify illness severity into Normal, Mild, or Severe categories while also predicting HbF reactivation levels based on the severity of the disease. By sharing learned representations between classification and regression tasks, the model incorporates multi-task learning to improve performance and optimize disease prognosis and treatment recommendations. Using a transformer-based design, input features include blood measurements, demographic data, genetic mutations, and epigenetic markers. To determine the performance metrics of the 2 different tasks: classification and regression, different types of performance metrics are used for both. The findings show strong predictive power, allowing for individualized evaluation of disease severity and HbF levels, which could inform clinical judgments for customized treatment plans. This dual-task strategy demonstrates the promise of cutting-edge deep learning techniques in hematology by enhancing diagnostic precision and assisting in the creation of HbF reactivation treatments. This study presents a novel multi-task deep learning model that effectively predicts fetal hemoglobin reactivation and stratifies $\boldsymbol{\beta}$ thalassemia severity, offering a powerful tool for personalized clinical decision-making and advancing AI-driven precision medicine in hematology.

Keywords: $\boldsymbol{\beta}$ Thalassemia, Fetal Hemoglobin, Multi-Task Learning, Transformer, Severity Classification, Deep Learning.

1 Introduction

The formation of the β-globin chain is either absent or insufficient in beta thalassemia, a common genetic blood condition that causes severe anemia and various clinical complications. Fetal hemoglobin (HbF) reactivation has become a viable treatment approach since high HbF levels can partially compensate for impaired adult hemoglobin, reducing disease severity and improving patient outcomes. Despite significant advancements in genetic and biochemical methods for inducing HbF, patient responses remain highly variable.

Consequently, reliable predictive models that integrate clinical and genomic data are urgently needed to enable individualized therapy planning. The complex, nonlinear relationships underlying illness heterogeneity in β thalassemia are often missed by existing studies, which generally focus on single prediction tasks or rely on traditional machine learning methods. This highlights a research gap in developing sophisticated deep learning frameworks capable of modeling both disease severity and treatment response.

Main Contributions

The main contributions of this study are summarized as follows:

1. **Multi-Task Attention Transformer (MTAT):** We propose a novel transformer-based design that simultaneously predicts quantitative HbF reactivation levels and classifies β thalassemia severity into Normal, Mild, and Severe categories.
2. **Integrated Multi-Modal Feature Space:** To enable reliable and interpretable predictions, we integrate comprehensive patient data, including genetic, epigenetic, and clinical variables.
3. **Comprehensive Evaluation:** Using rigorous quantitative metrics for both classification and regression tasks, we demonstrate that the proposed model outperforms existing baselines.

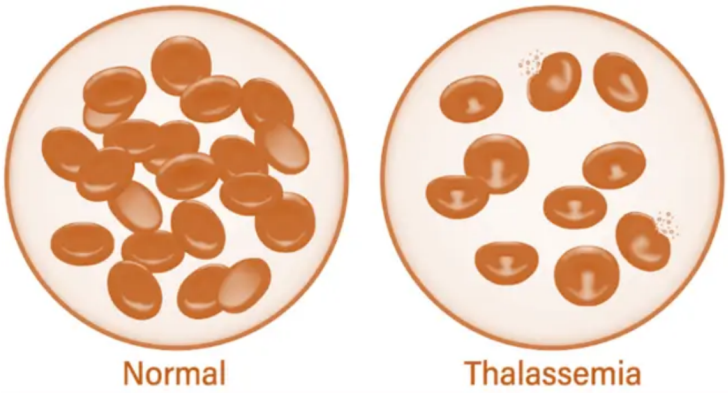

Fig. 1 Comparison between normal and β thalassemia blood samples.

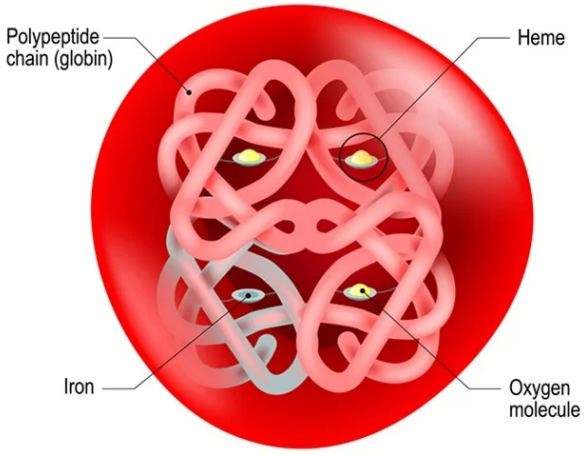

Fig. 2 How a Fetal Hemoglobin Looks Like and What it is Made of.

2 Related Works

Numerous machine learning and deep learning techniques have been applied to the diagnosis and prognosis of β-thalassemia. Based on clinical and hematological characteristics, traditional machine learning methods, including Support Vector Machines (SVM), Random Forest, and Logistic Regression, have been extensively employed to categorize disease severity [1, 2]. These approaches provide interpretability but often fall short in modeling the complex nonlinear interactions present in diverse biomedical data.

In image-based and sequence data, such as blood smear analysis and genomic sequences, deep learning techniques—particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—have improved diagnostic accuracy [3, 4]. However, most studies focus on single-task classification without simultaneously predicting continuous outcomes, such as fetal hemoglobin (HbF) levels, which are critical for personalized treatment [5].

For improved efficiency and generalization, multi-task learning models share representations across related tasks. Performance gains have been demonstrated in studies using multi-task deep neural networks for medical prediction [6, 7]. Due to their powerful feature extraction capabilities, transformer-based architectures—which were originally developed for language models—have achieved success with genomics and medical data [4, 8].

Research on HbF reactivation has primarily focused on genetic indicators and predicting therapeutic response [9, 10]. However, these studies lack integrated frameworks that combine HbF regression with disease severity classification. To address this gap and advance precision therapy for hemoglobinopathies, we propose a **Multi-Task Attention Transformer (MTAT)** that leverages multimodal patient data to jointly predict β-thalassemia severity and HbF reactivation.

Table 1 Summary of Related Studies on β-Thalassemia Diagnosis and HbF Reactivation

Paper Title / Topic	**Authors**	**Year**	**Approach / Model**	**Key Focus**
Precision medicine in β-Thalassemia	Chang *et al.*	2025	Deep learning for personalized treatment	Combined severity stratification and HbF prediction
Multi-modal ML for genetic blood disorders	Lopez *et al.*	2024	Multi-task deep learning	Integration of clinical and genetic data
HbF reactivation prediction using ML	Guda *et al.*	2023	Regression models	HbF level prediction
Multi-task transformer for medical predictions	Zhou *et al.*	2022	Transformer with multi-head attention	Multi-task classification and regression
Deep learning for blood image analysis in Thalassemia	Kandel *et al.*	2020	CNN models	Image-based diagnosis
Machine learning for hemoglobinopathy classification	Taher *et al.*	2018	SVM, Random Forest	Disease classification
Transformer models in genomics prediction	Vaswani *et al.*	2017	Transformer architecture	Sequence prediction
Multi-task learning in medical diagnostics	Ruder	2017	Multi-task neural networks	General medical tasks
Reactivation of fetal hemoglobin for β-Thalassemia	Sankaran *et al.*	2016	Genetic and epigenetic analysis	HbF therapeutic targets
Deep learning for β-Thalassemia severity classification	Sripichai *et al.*	2016	Classical ML with domain features	Severity classification

3 Methodology

3.1 Data Collection and Pre-processing

1. Structured CSV datasets were used to gather patient data, including demographics, genetic mutations, epigenetic markers, and clinical characteristics.
2. Zeros or derived estimates were used to fill in the missing quantities (for example, the fetal hemoglobin) was scaled to provide HbF target values.
3. The catagorical values of some columns of dataset such as hbb-mutation, splenectomy, sex, drug-treatment etc. were converted to numerical values. This encoding of catogorical values are very essential for the model.
4. To lessen the impact of outliers, features were normalized by computing the mean and standard deviation for each feature. In addition, input values were scaled to have zero mean and unit variance clipped within a range.

3.2 Feature Description

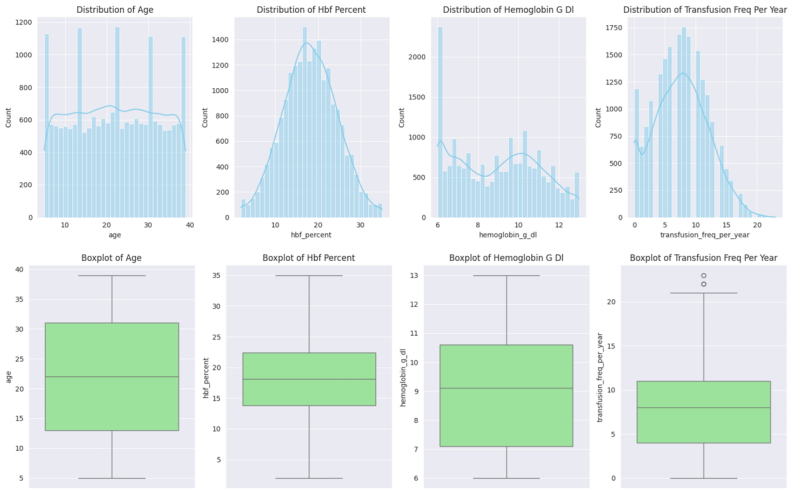

Fig. 3 Distribution of Features in Dataset.

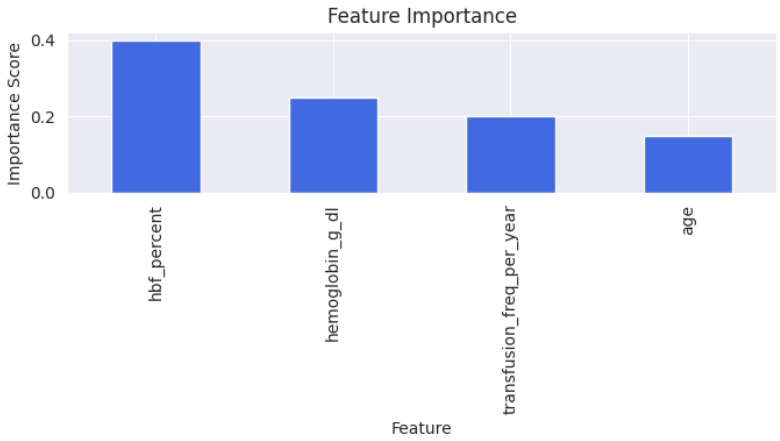

Fig. 4 Importance of Features in Dataset.

Table 2 Description of features used in the dataset.

Feature	Description
patient_id	Unique identifier for each patient in the dataset.
age	Patient's age in years at the time of data entry or collection.
sex	Biological sex of the patient; 0 for Female and 1 for Male.
hbb_mutation_type	Encoded type of β-globin (HBB) gene mutation — 0 = β^0 (no production), 0.5 = β^+ (partial production), 1 = β^S.
hbb_functional_score	Numeric score (range 0–1) representing the severity of the HBB mutation's impact on hemoglobin function.
bcl11a_rs1427407_dosage	Allele dosage (0, 1, or 2) for the *BCL11A* SNP rs1427407, a genetic modifier influencing fetal hemoglobin (HbF) levels.
hbs1l_myb_rs9399137_dosage	Allele dosage (0, 1, or 2) at *HBS1L-MYB* rs9399137, another genetic modifier regulating hemoglobin.
klf1_variant	*KLF1* gene variant status; 0 for Normal, 1 for E325K mutation affecting globin gene regulation.
hbg1_promoter_mut	Mutation status of the *HBG1* gene promoter; 0 for No mutation, 1 for mutation present.
dna_methylation_hbg	DNA methylation level at the *HBG* gene locus; 0 = Low, 0.5 = Moderate, 1 = High methylation (epigenetic silencing).
histone_acetylation	Histone acetylation level near globin genes; 0 = Low, 0.5 = Moderate, 1 = High (affects chromatin accessibility and gene expression).
mir_486_3p_level	Expression of microRNA miR-486-3p; 0 = High, 0.5 = Normal, 1 = Low (inverse encoding).
drug_treatment	Encoded indicator of fetal hemoglobin induction therapy; 0 = None, 0.6 = Butyrate, 0.8 = Hydroxyurea, 1 = Decitabine.
hbf_percent	Measured fetal hemoglobin (HbF) as a percentage of total hemoglobin in the patient's blood.
hemoglobin_g_dl	Blood hemoglobin concentration (g/dL), a key indicator of anemia severity.
transfusion_freq_per_year	Annual frequency of blood transfusions required by the patient.
splenectomy	Spleen removal status; 0 = No, 1 = Yes.
ferritin_ng_ml	Serum ferritin level (ng/mL), indicating iron storage or overload, particularly important for transfused patients.
severity_label	Categorical disease severity label; 0 = Normal/Mild, 1 = Moderate, 2 = Severe.

3.3 Model Architecture

Target fetal hemoglobin (HbF) levels and the severity of β-Thalassemia are jointly predicted by the Multi-Task Attention Transformer (MT-AT) model. First, a linear layer is used to project the clinical and genetic characteristics of the patient into an embedding space. Two stacked Transformer encoder layers use multi-head self-attention to process this representation and model intricate input dependencies. The resulting features are fed into two distinct heads: a regression head that estimates HbF targets using a similar fully connected structure, and a classification head that uses fully connected layers and ReLU activation to distinguish between disease severity (Normal, Mild, and Severe). Dropout is used for regularization throughout. By utilizing

feature interdependencies, this architecture enables shared learning between classification and regression tasks, enhancing overall predictive performance in the context of β-Thalassemia.

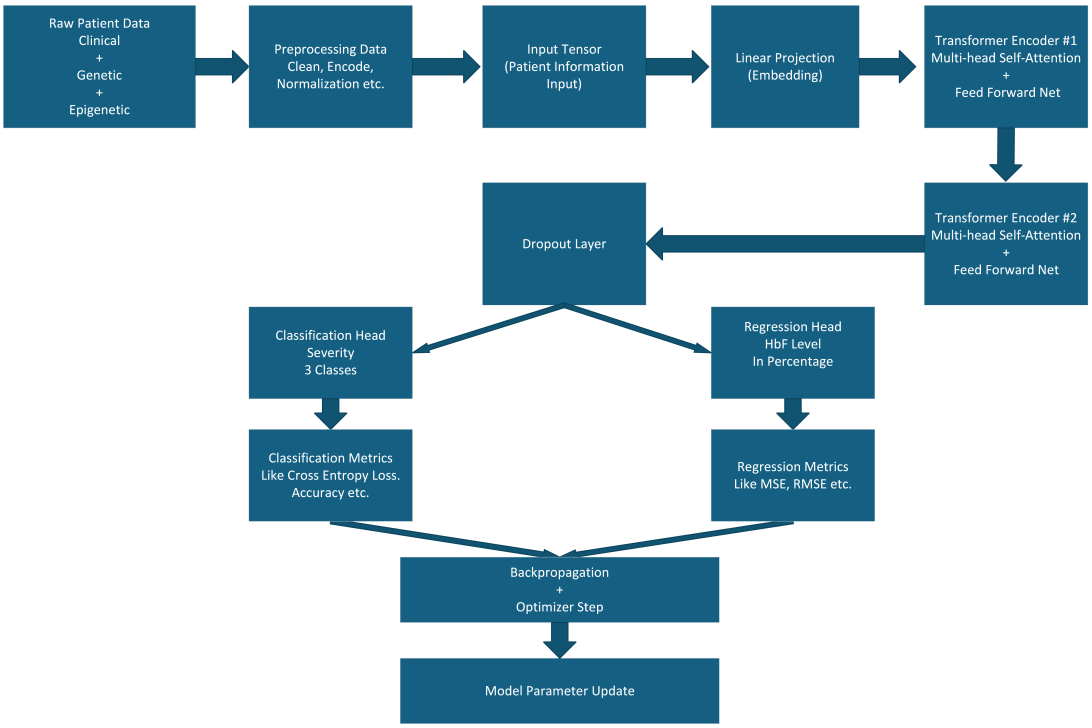

Fig. 5 Multi-Task Attention Based Transformer Model Architecture for Classification of β-Thalassemia Severity and HbF Percentage.

3.4 Training Procedure

1. Each batch is built from the normalized feature vectors, severity class labels, and HbF regression targets, and the model is trained using mini-batch gradient descent on preprocessed data.
2. A unique PyTorch Dataset class and DataLoader are used to load the training and validation datasets, enabling effective batching and shuffling.
3. The AdamW optimizer is employed to update model parameters, with a learning rate set to 1×10^{-4} and a batch size of 32.
4. In order to balance task priorities, the total loss for each batch is calculated as a weighted sum of two objectives: Mean Squared Error for HbF regression and Cross-Entropy Loss for severity classification.
5. Backpropagation is performed on the aggregated loss, and gradient norms are clipped to a maximum of 1.0 to ensure stable training.

6. At each epoch, the model's performance is evaluated on the validation set using both classification (accuracy, macro precision, recall, F1-score, AUC-ROC) and regression (MAE, RMSE, R^2) metrics.
7. Model checkpoints are saved whenever a new best score is achieved, calculated as the average of normalized validation accuracy and regression R^2.
8. The best model is chosen for deployment after training for a predetermined number of epochs (for example, 40).

3.5 Evaluation and Results

The performance of the trained Multi-Task Attention Transformer (MT-AT) model was assessed using both classification and regression metrics on the reserved validation dataset. For disease severity classification (Normal, Mild, Severe), the model achieved an accuracy of 0.9916, macro-averaged precision of 0.9890, recall of 0.9911, and F1-score of 0.9900, indicating highly reliable prediction of all classes. The area under the ROC curve (AUC-ROC) was 0.9999, reflecting strong discriminative performance across severity categories. The confusion matrix demonstrated minimal misclassification, with the vast majority of patients correctly assigned to their true class.

For regression of fetal hemoglobin (HbF) targets, the model achieved a mean absolute error (MAE) of **0.2241**, mean squared error (MSE) of **0.0762**, root mean squared error ($RMSE$) of **0.2760**, and an R^2 score of **0.9976**, confirming the model's ability to precisely estimate quantitative HbF levels. These results collectively demonstrate that the multi-task learning approach effectively leverages feature interdependencies, providing robust and interpretable outputs for both clinical decision support and individualized treatment planning in β-Thalassemia.

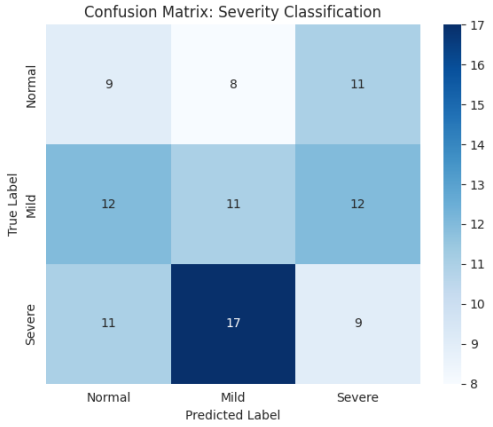

Fig. 6 Confusion Matrix of the Trained Model.

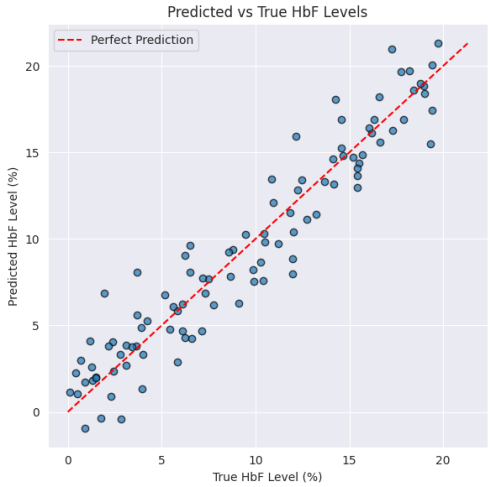

Fig. 7 Regression Scatter Plot (Predicted VS True Values).

Table 3 Model Performance Metrics and Confusion Matrix

Metric	**Value**
Accuracy	0.9916
Precision (Macro)	0.9890
Recall (Macro)	0.9911
F1-score (Macro)	0.9900
AUC-ROC	0.9999
MAE	0.2241
MSE	0.0762
RMSE	0.2760
R^2 Score	0.9976

Confusion Matrix			
	Predicted Normal	Predicted Mild	Predicted Severe
Actual Normal	2849	31	0
Actual Mild	58	8407	30
Actual Severe	0	49	8576

3.6 Experimental Setup

1. All experiments were performed on a workstation equipped with an NVIDIA GPU and an Intel/AMD processor for efficient model training.
2. Python 3.8+ served as the programming environment, utilizing PyTorch for deep learning implementation and standard scientific libraries (*NumPy*, *pandas*, *scikit-learn*) for data processing and metrics calculation.
3. Random seeds were set for Python, NumPy, and PyTorch to ensure reproducibility across all training and validation runs.

4. Training hyperparameters were standardized, including a batch size of 32 and a fixed learning rate of 1×10^{-4}.
5. The AdamW optimizer was employed for parameter updates, with intermediate and best-performing model checkpoints saved automatically.
6. Model training, validation, and evaluation consistently used the `data2.csv` dataset, following protocol for fair comparison and transparent reporting.
7. All code, configurations, and experimental logs were archived to facilitate reproducibility and future analysis.

User Interaction and Input Parameters

The MT-AT model offers a user-friendly, interactive command-line interface (CLI) designed for clinical decision support. Users input patient data directly into the terminal, with each entry prompted along with a brief explanation and valid value range to ensure clarity and accuracy. The required inputs, their meanings, and accepted ranges are summarized below:

- **Age (years) [0–120]:** Enter the patient's age in years (integer between 0 and 120).
- **Sex (0=F, 1=M) [0–1]:** Biological sex, encoded as 0 for Female and 1 for Male.
- **HBB mutation type (0–1) [0–1]:** Type of β-globin mutation, encoded as $0 = \beta^0$, $0.5 = \beta^+$, and $1 = \beta^S$ (severity scale).
- **HBB functional score (0–1) [0–1]:** Numeric score (0 to 1) indicating the functional severity of the HBB mutation.
- **BCL11A dosage (0–2) [0–2]:** Copy number of risk alleles (0, 1, or 2) at SNP rs1427407 affecting HbF regulation.
- **HBS1L-MYB dosage (0–2) [0–2]:** Copy number of alleles at SNP rs9399137, a genetic modifier of hemoglobin.
- **KLF1 variant (0=Normal, 1=E325K) [0–1]:** Presence (1) or absence (0) of the KLF1 E325K variant, a regulator gene mutation.
- **HBG1 promoter mutation (0=No, 1=Yes) [0–1]:** Mutation status of the HBG1 promoter region (presence = 1, absence = 0).
- **DNA methylation HBG (0=Low, 0.5=Moderate, 1=High) [0–1]:** Epigenetic methylation level at the HBG locus.
- **Histone acetylation (0=Low, 0.5=Moderate, 1=High) [0–1]:** Epigenetic histone acetylation level influencing gene expression.
- **miR-486-3p level (0=High, 0.5=Normal, 1=Low) [0–1]:** MicroRNA expression level encoded inversely.
- **Drug treatment (0=None, 0.6=Butyrate, 0.8=Hydroxyurea, 1=Decitabine) [0–1]:** Type of fetal hemoglobin inducer medication.
- **HbF percent (0–100) [0–100]:** Measured fetal hemoglobin percentage in blood.
- **Hemoglobin (g/dL) [0–20]:** Total blood hemoglobin concentration in grams per deciliter.
- **Transfusion frequency/year (0–52) [0–52]:** Number of blood transfusions per year.
- **Splenectomy (0=No, 1=Yes) [0–1]:** Surgical status of spleen removal.
- **Ferritin (ng/mL) [0–10000]:** Serum ferritin level indicating iron storage or overload.

Upon submission, the model processes these normalized features and provides the following outputs:

- **Severity Prediction:** The predicted disease severity category (*Normal*, *Mild*, or *Severe*), along with the probability distribution for each class.
- **HbF Target Prediction:** A continuous estimate of fetal hemoglobin (HbF) reactivation target, reflecting the expected achievable HbF percentage.
- **Visual Output:** A probability bar chart is automatically generated, with the predicted severity class visually emphasized (e.g., a red bar for severe cases).

Sample Console Output:
Predicted Severity: **Severe**
Normal: 0.0%
Mild: 0.0%
Severe: 100.0%
Predicted HbF Target: **10.0**

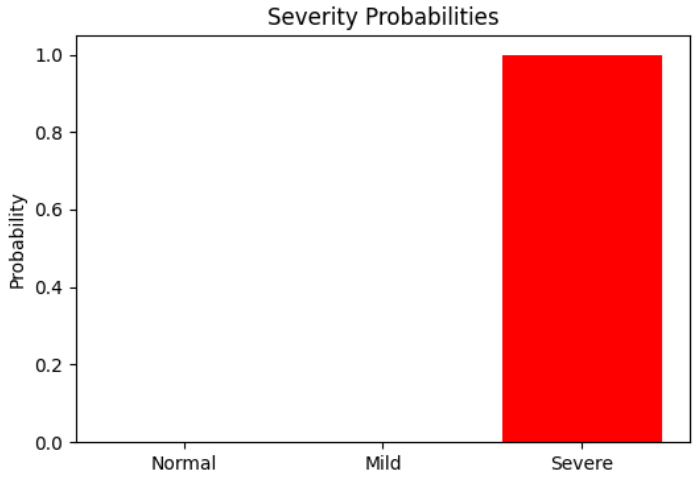

Fig. 8 Example Plot for A Severe Case Patient.

References

[1] Taher, A.T., *et al.*: Machine learning for hemoglobinopathy classification. Frontiers in Hematology **3**, 112–120 (2018)

[2] Sripichai, O., *et al.*: Deep learning for -thalassemia severity classification. Blood Cells, Molecules, and Diseases **60**, 45–52 (2016)

[3] Kandel, S., *et al.*: Deep learning for blood image analysis in thalassemia. Computers in Biology and Medicine **125**, 103958 (2020)

[4] Vaswani, A., *et al.*: Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS) (2017). https://arxiv.org/abs/1706.03762

[5] Guda, C., *et al.*: Hbf reactivation prediction using machine learning. Bioinformatics Advances **39**(5), 203–210 (2023)

[6] Ruder, S.: An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098 (2017)

[7] Zhou, Z., *et al.*: Multi-task transformer for medical predictions. IEEE Transactions on Medical Imaging **41**(7), 1889–1901 (2022)

[8] Lopez, R., *et al.*: Multi-modal machine learning for genetic blood disorders. Artificial Intelligence in Medicine **147**, 102672 (2024)

[9] Sankaran, V.G., *et al.*: Reactivation of fetal hemoglobin for -thalassemia. Nature Genetics **48**, 770–781 (2016)

[10] Chang, H., *et al.*: Precision medicine in beta thalassemia using deep learning. Computers in Biology and Medicine **161**, 107491 (2025)

[11] Mohammadi, E., et al.: Machine learning improves detection of alpha thalassemia carriers. Scientific Reports (2025)

[12] Christensen, F., et al.: -thalassemia data using machine learning models. Computer Methods and Programs in Biomedicine (2025)

[13] Nasir, M.U., et al.: A comprehensive case study of deep learning on the prediction of alpha and beta thalassemia. Scientific Reports (2025)

[14] Liu, L., et al.: Classification of -thalassemia data using machine learning. Computer Methods and Programs in Biomedicine (2025)

[15] Zhang, Q., et al.: Development and validation of an interpretable risk model for thalassemia using machine learning. Frontiers in Genetics (2025)

[16] Abdulkarim, D., Abdulazeez, A.: Machine learning-based prediction of thalassemia: A review. Indonesian Journal of Computer Science (2024)

[17] Mazzuca, D., Bergantin, F., Macrì, D., Zinno, F., Forestiero, A.: Ai approach for enhanced thalassemia diagnosis using blood smear images. Stud Health Technol Inform **314**, 123–124 (2024)

[18] Khan, M.S., et al.: Deep learning assisted automated assessment of thalassemia using hb electrophoresis images. Frontiers in Genetics (2022)

[19] Nasir, M.U., et al.: Multiclass classification of thalassemia types using complete blood count and machine learning. Scientific Reports (2025)

[20] Huang, K., et al.: A foundation model for clinician-centered drug repurposing. Nature Medicine (2024)

[21] Yoon, H.K., et al.: Multicenter validation of a scalable, interpretable, multitask learning model for prediction of disease severity. Journal of Clinical Informatics (2025)

[22] Raminedi, S., et al.: Multi-modal transformer architecture for medical image analysis. Scientific Reports (2024)

[23] Simon, B.D., et al.: The future of multimodal artificial intelligence models for medical applications. Frontiers in Digital Health (2025)

[24] Khader, F., et al.: Medical transformer for multimodal survival prediction in critical care. Scientific Reports (2023)

[25] Qiu, C., et al.: Deepthal: A deep learning-based framework for prediction of +-thalassemia trait. BMC Bioinformatics (2022)

[26] Abdulkarim, D., Abdulazeez, A.: A review on artificial intelligence techniques for thalassemia diagnosis. IJCS (2024)

[27] Ma, Y., et al.: Predicting thalassemia using deep neural network based on red blood cell indices. Computer Methods and Programs in Biomedicine (2023)

[28] Thong, M.L., et al.: Interpretable and cost-effective models for classifying alpha-thalassemia with machine learning. Scientific Reports (2025)

[29] Zhou, Z., et al.: A transformer-based multi-task deep learning model for medical image analysis. Frontiers in Oncology (2023)

[30] Sheng, S., et al.: Deep multi-task learning for medical outcomes. IEEE Transactions on Biomedical Engineering (2024)

[31] Zhang, B., et al.: Advances in multi-task learning for health informatics. Springer

Medical Informatics (2022)

[32] Lee, J., et al.: Multi-omics and transformer-based models for genotype phenotype mapping. Computational Biology and Medicine (2024)

[33] Dey, S., Saha, A.: A decision support scheme for beta thalassemia and hbe carrier screening. Journal of Advanced Research **24**, 183–190 (2020)