HBV-DS: Hepatitis B Virus Dataset for Predicting Liver Fibrosis and Viral Activity Using Machine Learning

Hanen Akkari
Laboratory of Advanced Technology
Intelligent Systems
National School of Engineering Sousse
University of Sousse, Tunisia
Email: hanen.bettaieb@gmail.com

Imen Akkari
Department of Gastroenterology
University Hospital of Hached,
University of Sousse,
Faculty of Medicine of Sousse, Tunisia
Email: imenakkaribm@gmail.com

Salma Haj Salah
Laboratory of Advanced Technology
Intelligent Systems
National School of Engineering Sousse
University of Sousse, Tunisia

Jinen Daghrir
Laboratory SIME, ENSIT,
University of Tunis
Tunis, Tunisia
Email: jinen.daghrir@isitc.u-sousse.tn
0000-0002-1300-8939

***Abstract*—This study introduces the HBV-DS dataset, a novel clinical resource designed for predicting hepatic fibrosis and necroinflammatory activity in patients with Hepatitis B Virus (HBV). The dataset is unique as it comprises the same patient instances annotated with multiple classification schemes, allowing for a comprehensive analysis of liver conditions. We implemented a robust end-to-end machine learning (ML) pipeline that includes essential preprocessing steps such as data cleaning, normalization, and encoding, alongside class balancing using the Synthetic Minority Oversampling Technique (SMOTE). This approach effectively addresses class imbalance, which is critical for enhancing model performance in clinical settings. Our evaluation involved six different classifiers: Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and XGBoost. Notably, the Random Forest classifier achieved an impressive accuracy of 92% and an AUC of 0.98, demonstrating the dataset's effectiveness in clinical prediction tasks. The findings highlight the significant impact of dataset-specific labeling on predictive outcomes and establish benchmark metrics for future research in HBV progression modeling. This work not only underscores the importance of preprocessing and model selection in medical machine learning applications but also provides actionable insights for integrating these models into clinical decision support systems.**

***Index Terms*—Hepatitis B, fibrosis, activity, machine learning, SMOTE, feature selection, medical data preprocessing.**

I. Introduction

HBV remains a major global health challenge, affecting approximately 296 million people and significantly contributing to liver-related morbidity and mortality [1]. An accurate liver fibrosis and viral activity assessment guides clinical decisions and optimizes patient management. However, conventional diagnostic methods, such as liver biopsy, are invasive, costly, and not widely accessible. Noninvasive approaches, including serum biomarkers and elastography, have been developed, but their accuracy remains suboptimal [2]. Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for analyzing complex biomedical data and predicting disease progression [3]. In the context of HBV, supervised learning models have been explored to estimate liver fibrosis progression. However, their performance critically depends on three key factors: (1) the quality and representativeness of the dataset, (2) the discriminative power of the selected clinical and biological features, and (3) the robustness of the chosen model in capturing underlying patterns. To address these challenges, we propose a rigorously curated dataset paired with an optimized ML framework to improve the accuracy and generalizability of HBV fibrosis prediction. This paper presents a comprehensive machine learning framework for predicting liver fibrosis and HBV activity using one of the four distinct clinical datasets. Each dataset is targeting specific diagnostic categorizations: (1) significant liver lesions (BD-Lesion), (2) hepatic activity (BD-Activity), (3) fibrosis staging (BD-Fibrose), and (4) combined fibrosis-activity assessment (HBV-DS). Our supervised learning approach combines rigorous feature selection with advanced classification models, optimized through stratified cross-validation techniques to ensure clinical relevance. The study makes three key contributions: First, we demonstrate how dataset-specific labeling impacts predictive performance across six machine learning classifiers. Second, we establish benchmark metrics (precision, sensitivity, specificity, and AUC-ROC) for each diagnostic task. Third, we provide actionable insights for integrating these models into clinical decision support systems. This paper is structured as follows: Section II presents a review of existing methods for predicting fibrosis and HBV activity. Section III describes

the proposed methodology, including data description, data selection, pre-processing, and model implementation. Section IV presents the experimental results and their interpretation, followed by a discussion of clinical implications and potential improvements in Section V. Finally, Section VI concludes with a summary of contributions and future research directions.

II. Related Work

Non-invasive assessment of liver fibrosis in chronic hepatitis B (CHB) has evolved significantly over the past decade. The gold standard liver biopsy presents inherent limitations, including sampling error, invasiveness, and complications, driving research toward alternative diagnostic approaches. Marcellin et al. [8] initially validated transient elastography for fibrosis assessment in CHB patients, demonstrating AUROCs of 0.81 and 0.93 for significant fibrosis and cirrhosis, respectively. Serum biomarker panels emerged concurrently, with Poynard et al. [9] developing FibroTest-ActiTest, achieving sensitivities of 70-86% for significant fibrosis detection. Despite these advances, these singular approaches showed limitations, particularly in intermediate fibrosis stages [10]. Recently, machine learning has transformed this landscape considerably. Jiang et al. [11] developed both nomogram and random forest models for CHB patients that achieved remarkable discrimination for significant fibrosis (AUROCs of 0.891 and 0.906, respectively) using only routine clinical variables. Anteby et al. [12] conducted a comprehensive systematic review of deep learning applications for fibrosis classification, identifying consistent superiority of AI techniques over conventional methods across multiple studies. Integration of multimodal data has further improved diagnostic performance, as demonstrated by Wang et al. [13], who combined radiomics features from MRI with clinical parameters through deep learning architectures. Despite these technological advances, Zhang et al. [14] identified persistent challenges in model generalization across diverse populations, emphasizing the need for external validation and population-specific calibration. The combined approach utilizing sequential algorithms of serum markers and elastography, as validated by Liu et al. [15], currently represents the most robust clinical strategy for non-invasive fibrosis assessment in CHB. Kim et al. [16] further refined machine learning applications by identifying novel biomarker combinations with improved prognostic value beyond fibrosis staging alone. While these recent advances in non-invasive fibrosis assessment and machine learning applications for chronic hepatitis B (CHB) have shown promising outcomes, these existing approaches often focus on either fibrosis or inflammatory activity in isolation. This limitation highlights the need for standardized, reproducible pipelines that are adapted to structured, multi-label clinical datasets. Moreover, few studies adequately address the challenges of data imbalance and the necessity for population-specific validation. In response to these limitations, our work introduces an integrated and reproducible end-to-end machine learning (ML) framework tailored to a curated HBV dataset. This framework targets the simultaneous prediction of fibrosis stage and inflammatory activity, thereby filling a critical gap in the current research landscape. By supporting clinically relevant, population-specific modeling in HBV management, our contribution aims to enhance the accuracy and applicability of predictive models in real-world clinical settings. This approach not only improves diagnostic capabilities but also aligns with the ongoing efforts to refine non-invasive assessment methods for liver conditions associated with HBV.

III. Methods

A. Dataset Description

This clinical dataset is derived from a retrospective investigation conducted at Farhat Hached University Hospital in Sousse, Tunisia. The collection comprises carefully curated medical records from 69 patients diagnosed with chronic hepatitis B (CHB). All cases met stringent diagnostic criteria for CHB and were selected based on complete clinical and biological data availability. The dataset's particular strength lies in its dual annotation system, supporting both binary classification (e.g., significant hepatic lesions) and multi-class staging (e.g., METAVIR fibrosis scoring) derived from a single patient group. A total of 30 features were collected for each patient, encompassing demographic, clinical, biochemical, and histological data. The full list of features is presented in Table I. Hepatitis B infection was confirmed through standard serological testing. The staging of liver fibrosis and grading of necroinflammatory activity were determined by histological examination of liver biopsy samples, using established scoring systems such as METAVIR. Evaluating the severity of fibrosis and activity is clinically crucial, as therapeutic decisions—particularly the initiation of antiviral treatment—are guided by these parameters. According to clinical guidelines, treatment is typically indicated only in cases of significant hepatic lesions, defined as an activity grade of A2 or higher and/or a fibrosis stage of F2 or higher. We have derived four specialized datasets from the original clinical records, each designed for distinct classification tasks. For the present analysis, we focus exclusively on the fourth dataset (HBV-DS), which enables four-class classification based on the combined clinical significance of fibrosis stages and necroinflammatory activity levels, as detailed in Table II. This curated dataset provides the most comprehensive representation of disease progression, aligning with standard histopathological grading systems while optimizing predictive performance for clinical decision-making.

B. End-to-end ML framework

The proposed methodology for predicting fibrosis and necroinflammatory activity in chronic hepatitis B follows a multi-step machine learning pipeline, as illustrated in Figure 1. Our pipeline begins with data preprocessing, handling missing values, encoding categorical variables, normalizing features, and balancing class distributions. Next, feature selection employs three methods: RFE-CV, SelectKBest, and Inf-FS, to identify optimal feature subsets. Six classifiers (MLP, SVM, KNN, XGBoost, Decision Tree, and Random Forest) are then trained and evaluated using both full and reduced feature

TABLE I
DESCRIPTION OF THE DATASET FEATURES

Feature Name	Description	Type
Age	Patient's age (in years)	Continuous
Gender	Male or Female	Categorical
Weight	Weight in kilograms	Continuous
Height	Height in centimeters	Continuous
BMI	Body Mass Index	Continuous
ALT	Alanine transaminase	Continuous
AST	Aspartate transaminase	Continuous
Platelets	Platelet count (10^9/L)	Continuous
Prothrombin Rate	Coagulation profile (%)	Continuous
INR	International Normalized Ratio (blood-clotting test)	Continuous
Albumin	Serum albumin level	Continuous
Total Bilirubin	Total bilirubin in blood	Continuous
Direct Bilirubin	Direct bilirubin level	Continuous
ALP	Alkaline phosphatase	Continuous
GGT	Gamma-glutamyl transferase	Continuous
HBV DNA	Viral load (copies/mL)	Continuous
HBeAg	Hepatitis B e-antigen	Binary
Anti-HBe	Anti-HBe antibody status	Binary
Diabetes Mellitus	Presence of diabetes	Binary
Hypertension	Presence of hypertension	Binary
Alcohol Use	Alcohol consumption	Categorical
Smoking	Smoking status	Categorical
Family History	Family history of liver disease	Binary
Previous Treatment	Antiviral treatment history	Binary
Liver Stiffness	From elastography or biopsy	Continuous
Biopsy Performed	Whether a biopsy was done	Binary
Fibrosis Score	METAVIR F0–F4 score	Ordinal / Binary
Activity Score	METAVIR A0–A3 grade	Ordinal / Binary
Diagnosis Date	Date of diagnosis	Date
Follow-up Period	Duration of monitoring	Continuous

TABLE II
MULTI-CLASS CLASSIFICATION OF LIVER ACTIVITIES AND FIBROSIS

Classe	Signification
0	Non-significant activity and fibrosis
1	Significant activity and non-significant fibrosis
2	Significant activity and fibrosis
3	Significant activity and advanced fibrosis

sets. Performance is assessed via metrics (accuracy, precision, recall, F1-score, and AUC) to compare model effectiveness in predicting HBV fibrosis and activity stages.

C. Data Preprocessing

The HBV-DS dataset presented specific modeling challenges, including missing values and class imbalance. Without proper handling, these issues could compromise model accuracy, fairness, and generalization. Our preprocessing pipeline addressed these limitations by transforming raw clinical data into a machine-learning-ready format while preserving data integrity. Key steps included missing value imputation, categorical encoding, feature normalization, and SMOTE-based class balancing. Each component is detailed in subsequent sections.

1) Missing data imputation:

- Imputation of Missing Numerical Values

To handle missing values in the numerical variables of the HBV-DS dataset, we experimented with three distinct imputation techniques. These included mean imputation, median imputation [4], [5], and the K-Nearest Neighbors (KNN) imputation method. Each approach was evaluated for its ability to preserve the underlying data distribution and minimize distortion before training the predictive models.

- Imputation of Missing Categorical Values

For categorical variables, a different strategy was employed to address missing values. Specifically, we applied mode imputation, replacing missing entries with the most frequently occurring value for each variable. This approach is particularly suitable for categorical data, where the most common category often provides a reliable estimate that reflects the overall distribution.

2) Encoding of Categorical Variables: We used label encoding (ordinal encoding) to convert categorical variables into a numerical format. This method assigns ascending integer values to each unique category, then replaces the original categorical values with these numerical labels, maintaining the variable's categorical meaning in a machine-readable form.

3) Data Normalization: We applied standardization (z-score normalization) to address scale variations among features. This critical preprocessing step transforms each feature to have a mean of 0 and a standard deviation of 1, preventing variables with larger ranges from dominating the model's learning process.

4) Synthetic Minority Over-sampling Technique (SMOTE): During our analysis, we observed that the HBV-DS dataset suffers from class imbalance, with certain categories being significantly underrepresented. Specifically, 50.24% of the

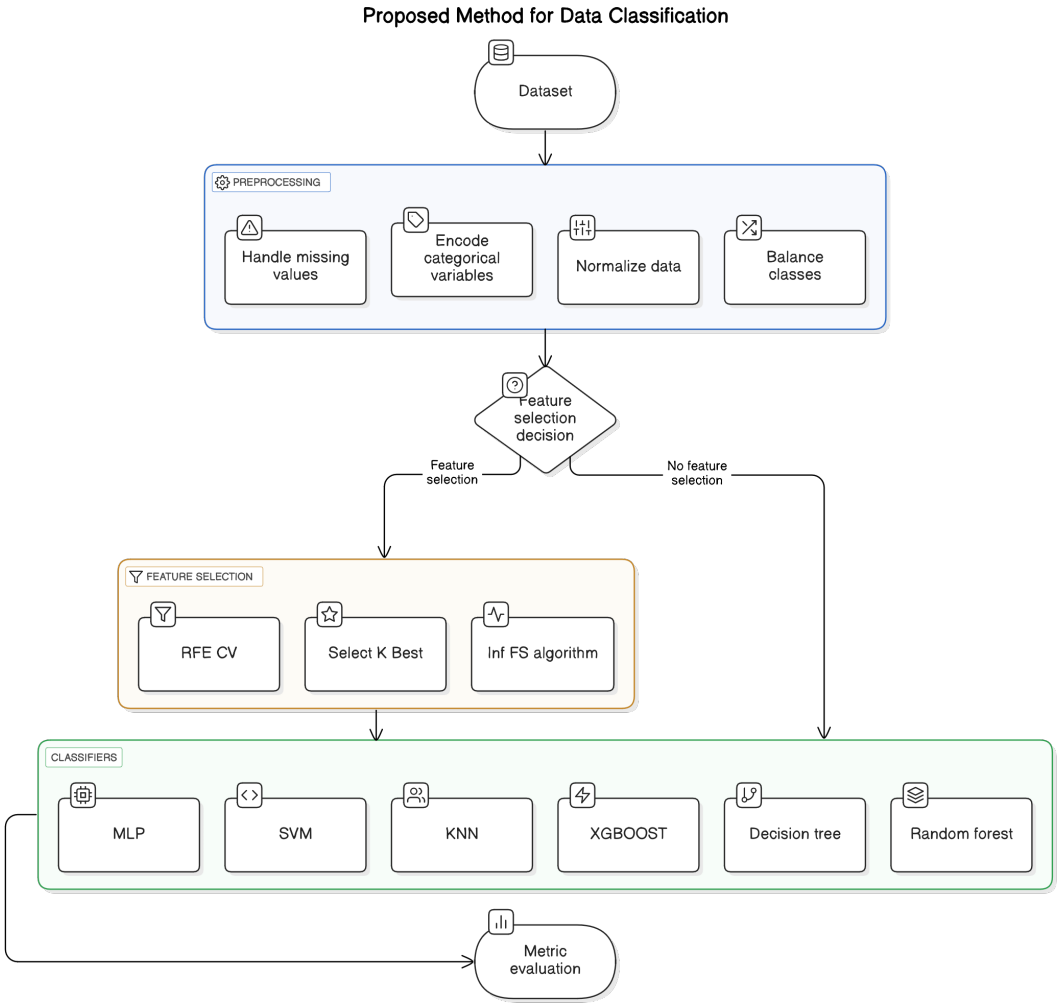

Fig. 1. Flowchart of the proposed solution

data belong to Class 0 (non-significant activity and fibrosis), 21.74% to Class 1 (significant activity with non-significant fibrosis), 13.04% to Class 2 (significant activity and significant fibrosis), and only 5.8% to Class 3 (significant activity with advanced fibrosis). This imbalance poses a serious challenge, as machine learning models tend to favor the majority classes, leading to poor classification performance on minority classes. To address this issue, we adopted SMOTE (Synthetic Minority Over-sampling Technique), a widely used and effective re-sampling method introduced in 2002 [6]. Rather than simply duplicating existing minority instances, SMOTE generates synthetic samples by interpolating between selected instances and their nearest neighbors within the same class. This approach helps balance the class distribution and enhances the model's ability to learn from underrepresented categories. Following Yağanoğlu's [7] successful SMOTE application in hepatitis C research, we implemented this technique for our hepatitis B dataset to address class imbalance. This advanced resampling approach enhanced model fairness, accuracy, and stability, crucial requirements for clinical prediction tasks.

D. Feature Selection

Feature selection is a critical step in the ML framework, aiming to identify and retain the most informative variables for the prediction task. By eliminating redundant, irrelevant, or noisy features, this step not only enhances model performance but also reduces computational complexity, leading to better generalization on unseen data. In the context of this study, we sought to determine which clinical and biological variables are most relevant for predicting fibrosis stage and hepatic activity in patients with chronic hepatitis B. To that end, we evaluated and compared three feature selection techniques:

- Recursive Feature Elimination with Cross-Validation (RFE-CV),
- SelectKBest, which selects features based on univariate statistical tests,
- Infinite Feature Selection (Inf-FS), a graph-based ranking algorithm.

These methods were applied to the HBV-DS dataset to isolate the most predictive features, to improve the model's interpretability and diagnostic performance in a clinical context.

E. Machine Learning Models

In this study, we implemented and compared several supervised Machine Learning models to classify patients based on the severity of hepatic fibrosis and necroinflammatory activity. The models were chosen for their widespread use in medical data analysis and their diversity in learning approaches. The configurations are summarized below:

1) K-Nearest Neighbors (KNN): We employed this distance-based, non-parametric algorithm, which classifies patients by comparing each new instance to its K closest neighbors in the training data (using Euclidean distance). Model performance was evaluated across K values ranging from 3 to 15.

2) Support Vector Machine (SVM): We implemented this margin-maximizing algorithm using both linear and non-linear kernels (including RBF). The hyperparameters such as the regularization, and gamma, were optimized via grid search to balance model complexity and generalization for fibrosis/activity classification.

3) Multilayer Perceptron (MLP): MLP is a class of feedforward artificial neural networks that has become fundamental for solving complex classification problems. In this study, we implemented an MLP with the following architecture: a single hidden layer containing 100 neurons, using ReLU activation functions and the Adam optimizer. The model was trained for a maximum of 200 epochs with early stopping enabled to prevent overfitting. This configuration provided an optimal balance between model complexity and computational efficiency for our clinical classification task of hepatic fibrosis and inflammatory activity, while ensuring reproducible results through controlled randomization.

4) Decision Tree (DT): DT represents a powerful supervised learning model particularly effective for solving classification problems. Its fundamental principle relies on a tree-like structure, where each leaf node corresponds to a class label, while internal nodes represent decision rules derived from data features. This architecture offers inherent interpretability, allowing for transparent decision processes through sequential feature evaluations that culminate in classification outcomes.

5) Random Forest (RF): Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode (classification) or mean (regression) of their predictions. For our study, we implemented a Random Forest classifier composed of 100 decision trees, leveraging their collective decision-making to improve prediction accuracy and robustness while naturally handling overfitting through built-in feature randomness and bootstrap aggregation. This configuration provided optimal performance for our hepatic fibrosis classification task.

6) XGBoost: XGBoost (eXtreme Gradient Boosting) is a powerful and efficient implementation of gradient boosting tailored for supervised learning tasks. It builds decision trees sequentially, with each tree correcting the residuals of the previous one, thereby improving predictive performance. Known for its scalability and built-in regularization, XGBoost handles missing data, supports parallel and distributed computing, and employs level-wise tree growth for effective optimization. Key hyperparameters influencing performance include the learning rate (**learning_rate**), number of boosting rounds (**n_estimators**), tree depth (**max_depth**), and regularization terms (**reg_alpha**, **reg_lambda**). Proper tuning of these parameters is essential to balance accuracy and efficiency. To ensure robust and unbiased evaluation, all models were validated using 10-fold cross-validation. This method divides the dataset into ten subsets, training the model on nine and validating on the remaining one, rotating through all folds. The final performance is averaged over the ten iterations, providing a reliable estimate of model generalization. This consistent validation strategy was applied across all models to ensure fair and comparative assessment.

TABLE III
PERFORMANCE COMPARISON OF IMPUTATION METHODS ACROSS MODELS

Imputation Method	**Metric**	**MLP**	**KNN**	**SVM**	**RF**	**DT**	**XGB**
Numeric: Mean	Accuracy	0.60	0.63	0.63	0.63	0.69	0.72
Categorical: Mode	Precision	0.50	0.50	0.42	0.30	0.54	0.57
	Recall	0.40	0.32	0.36	0.40	0.39	0.45
	F1-Score	0.45	0.40	0.30	0.30	0.39	0.50
Numeric: Median	Accuracy	0.63	0.63	0.63	0.63	0.67	0.69
Categorical: Mode	Precision	0.55	0.53	0.40	0.45	0.50	0.45
	Recall	0.45	0.40	0.40	0.45	0.50	0.40
	F1-Score	0.40	0.42	0.35	0.25	0.40	0.25
Numeric: KNN (k=9)	Accuracy	**0.74**	0.63	0.50	0.63	0.67	0.71
Categorical: Mode	Precision	0.60	0.53	0.50	0.45	0.50	0.55
	Recall	0.55	0.46	0.45	0.45	0.50	0.60
	F1-Score	0.45	0.45	0.43	0.45	0.40	0.55

IV. RESULTS AND DISCUSSION

This section presents the outcomes of experiments conducted on the HBV-DS dataset, aimed at predicting fibrosis and activity in chronic hepatitis B. The experiments were performed using Python 3.10, using different libraries such as scikit-learn, pandas, etc, for model implementation, preprocessing, and evaluation. All tests were executed on a machine equipped with 16 GB of RAM and an Intel Core i7 processor.

A. Impact of Missing Data Imputation

To handle missing values, we conducted a comparative analysis of three imputation strategies for numerical features: mean, median, and K-Nearest Neighbors (KNN) with $k = 9$. For categorical features, the most frequent (mode) value was used across all approaches. As shown in Table III, KNN imputation with $k = 9$ consistently achieved superior performance, effectively preserving the underlying data distribution while minimizing bias. Notably, the impact of each imputation method varied across models, highlighting that certain strategies are better suited to specific learning algorithms.

B. Effect of SMOTE Oversampling

To mitigate class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was employed, resulting in an equal class distribution of 25% per class. The effectiveness of this approach is demonstrated in Table IV, where notable performance gains were observed across all models. In particular, MLP, XGBoost, and Random Forest (RF) exhibited substantial improvements in recall and F1-score, underscoring SMOTE's positive impact on model sensitivity and overall classification quality.

C. Feature Selection

To reduce redundancy and enhance interpretability, three feature selection techniques were applied: RFE-CV, SelectKBest, and Inf-FS. For RFE-CV, the optimal subset consisted of 21 features, among which key variables such as BMI, ALT, AST, and INR were retained. With SelectKBest, features were ranked based on their univariate statistical relevance, and those with a score greater than or equal to 5 were selected. This resulted in 13 top features, with CVinitiale (initial viral load), ALAT, and platelet count (PLAQ) emerging as the most significant. Similarly, Inf-FS ranked features by importance weights, and 16 features with weights 1 were retained. The top-ranked feature corresponded to index 3 in the dataset. Overall, these selection methods consistently highlighted clinically relevant variables, forming a robust foundation for downstream classification tasks.

1) Comparative Analysis of Feature Selection Methods: The impact of feature selection on classification performance was evaluated using three methods: RFE-CV (21 features), SelectKBest (13 features), and Inf-FS (16 features). As shown in Table V, RF and MLP consistently outperform other models across all feature selection techniques. RF achieves the highest performance with Inf-FS, recording an accuracy and F1-score of 0.90. MLP follows closely, reaching an accuracy of 0.88 and an F1-score of 0.86 with RFE-CV, and an F1-score of 0.85 with Inf-FS. In contrast, SelectKBest results in lower performance, with no model exceeding an F1-score of 0.79. SVM, KNN, and DT yield suboptimal results, with F1-scores remaining below 0.85 across all selection methods. Overall, Inf-FS demonstrates superior effectiveness by selecting a compact and highly discriminative feature subset and is thus selected for use in subsequent experiments.

D. Final Classification Results of the Proposed Solution

The proposed solution integrates optimized preprocessing steps using the most effective methods identified in prior experiments to predict fibrosis and activity in chronic hepatitis B using the HBV-DS dataset. Missing values were handled using KNN imputation ($K = 9$) for numerical features and mode imputation for categorical features. Class imbalance was addressed with SMOTE, achieving a balanced 25% distribution for each class. Features were selected using Inf-FS, retaining 16 highly representative and discriminative features. The models were trained with hyperparameters optimized via GridSearchCV using 5-fold cross-validation. Random Forest (RF, max_depth=10, n_estimators=200) and Multilayer Perceptron (MLP, hidden_layer_sizes=(50,50,50), activation=relu, solver=adam) emerged as the top-performing models. Table VI presents the final classification results. RF achieved the highest performance, with an accuracy of 0.92, precision of 0.91, recall of 0.91, F1-score of 0.90, and AUC of 0.98. MLP followed closely with an accuracy of 0.88 and an F1-score of 0.85. The confusion matrix and AUC curve in Figure **??** further confirm RF's robustness, with True Positives of 33, 39, 38, and 41 for classes 0 through 3, respectively.

The confusion matrix in Figure **??** confirms the RF model's ability to accurately classify all classes. These results highlight the strong performance of RF, SVM, and MLP, with RF demonstrating the highest overall accuracy and AUC. This underscores its robust generalization capabilities and effectiveness in predicting fibrosis and inflammatory activity in patients with chronic hepatitis B. The findings of this study emphasize the importance of integrating robust preprocessing techniques such as SMOTE for class balancing and Inf-FS for feature selection, with advanced ML models. The consistent outperformance of RF and MLP across all evaluation metrics is likely attributable to their ability to model complex, non-linear relationships in clinical data. The use of GridSearchCV and cross-validation further contributed to robust generalization despite the dataset's modest size. Nonetheless, some limitations must be acknowledged. The relatively small sample size may constrain the generalizability of the results across broader patient populations. Additionally, while SMOTE effectively addresses class imbalance, the synthetic samples may not fully reflect the clinical variability of real-world cases. Future work should explore integrating imaging data, longitudinal information, or additional clinical features to enhance the pipeline's predictive accuracy and utility in personalized clinical decision support.

V. Conclusion

This study introduced the HBV-BD dataset, a curated clinical dataset for assessing hepatic fibrosis and necroinflammatory activity in patients with chronic hepatitis B. To evaluate its usability and clinical relevance, we proposed an end-to-end machine learning pipeline encompassing robust data preprocessing, including missing value imputation, feature normalization, class balancing using SMOTE, and comparative feature selection techniques. Six machine learning models were developed and assessed across four clinical datasets. In conclusion, the proposed HBV-DS dataset and ML framework demonstrate both strong predictive performance and clinical relevance. Future work will focus on external validation and multi-modal data integration to further improve generalization and clinical impact. The HBV-BD dataset introduced in this work aims to support future research in predictive modeling for chronic hepatitis B. It can be shared privately upon request by contacting Ms. Imen Akkari by filling out this form https://docs.google.com/forms/d/e/1FAIpQLScGp3f5W_XHngUN57ztdDOGJab9_lQMioDBcBxNjlEfPWEPvA/viewform?usp=dialog.

TABLE IV
PERFORMANCE COMPARISON BEFORE AND AFTER SMOTE

Model	Without SMOTE				With SMOTE			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
MLP	0.42	0.21	0.17	0.15	0.75	0.78	0.75	0.71
SVM	0.58	0.50	0.49	0.49	0.57	0.48	0.47	0.44
KNN	0.49	0.26	0.22	0.24	0.49	0.42	0.37	0.33
XGBoost	0.48	0.50	0.13	0.21	0.70	0.60	0.55	0.55
RF	0.46	0.36	0.31	0.37	0.69	0.56	0.58	0.52
DT	0.26	0.15	0.08	0.10	0.55	0.47	0.39	0.38

TABLE V
MODEL PERFORMANCE WITH DIFFERENT FEATURE SELECTION METHODS

Model	RFE-CV (21 Features)				SelectKBest (13 Features)				Inf-FS (16 Features)			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
MLP	0.88	0.88	0.87	0.86	0.80	0.79	0.79	0.79	0.84	0.90	0.81	0.85
SVM	0.86	0.86	0.86	0.85	0.75	0.77	0.75	0.74	0.82	0.83	0.82	0.81
KNN	0.73	0.74	0.73	0.70	0.70	0.74	0.74	0.74	0.71	0.72	0.72	0.69
DT	0.75	0.74	0.73	0.70	0.76	0.77	0.77	0.74	0.74	0.76	0.77	0.74

TABLE VI
FINAL MODEL PERFORMANCE ON THE HBV-DS DATASET

Model	**Accuracy**	**Precision (per class)**				**Recall (per class)**				**AUC**
		0	1	2	3	0	1	2	3	
MLP	0.88	0.93	0.80	0.89	0.93	0.61	0.95	0.95	1.00	0.98
SVM	0.91	0.94	0.83	0.98	0.93	0.76	0.93	0.98	1.00	0.98
KNN	0.83	0.95	0.71	0.95	0.80	0.44	0.90	0.98	1.00	0.94
DT	0.77	0.60	0.71	0.89	0.93	0.76	0.61	0.80	0.90	0.87
RF	**0.92**	0.89	0.89	0.93	0.98	0.80	0.95	0.93	1.00	**0.98**
XGBoost	0.85	0.76	0.76	0.95	0.93	0.68	0.85	0.90	0.95	0.97

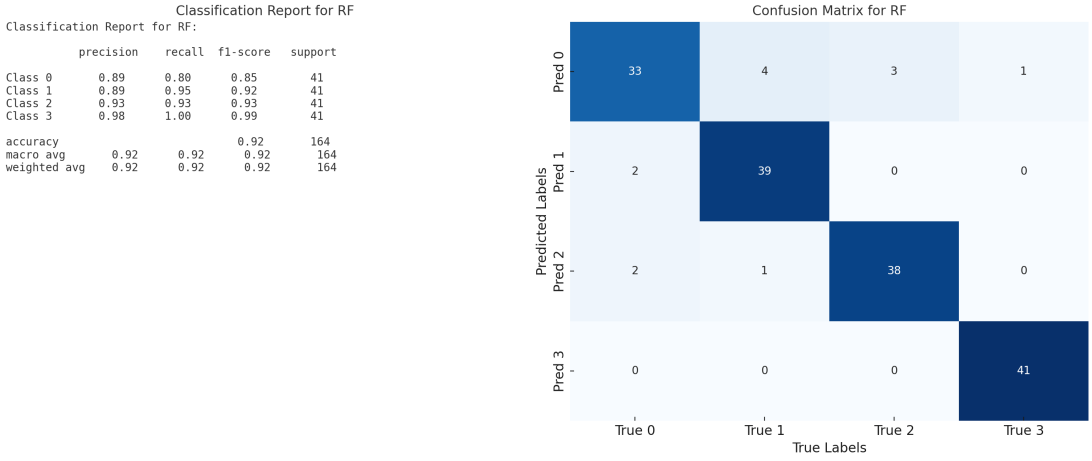

Fig. 2. Confusion matrix of the Random Forest model

ACKNOWLEDGEMENTS

The authors would like to sincerely thank the entire teams of the Gastroenterology and Infectious Diseases Departments at Farhat Hached Hospital, Sousse, Tunisia, for their invaluable support and collaboration in the collection and annotation of the dataset.

REFERENCES

[1] Bihl, F., Negro, F. (2009). Chronic hepatitis E in the immunosuppressed: a new source of trouble?. Journal of hepatology, 50(2), 435-437.

[2] Antona, D., Couturier, E., Larsen, C. (2011). Épidémiologie des hépatites virales en France: HEPATITES VIRALES. La Revue du praticien (Paris), 61(1).

[3] European Association For The Study Of The Liver. (2012). EASL clinical practice guidelines: management of chronic hepatitis B virus infection. Journal of hepatology, 57(1), 167-185.

[4] Schafer, J. L., Graham, J. W. (2002). Missing data: our view of the state of the art. Psychological methods, 7(2), 147.

[5] Van Buuren, S., Van Buuren, S. (2012). Flexible imputation of missing data (Vol. 10, p. b1182). Boca Raton, FL: CRC press.

[6] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

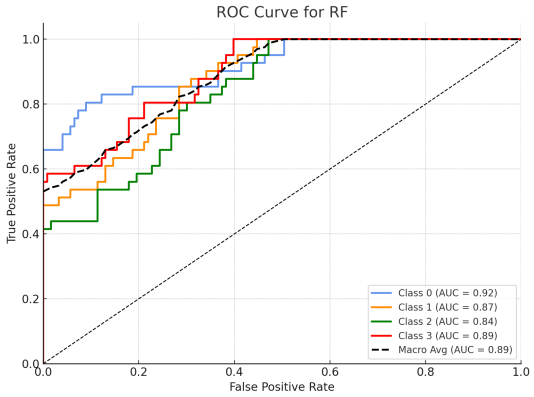

Fig. 3. AUC curve of the Random Forest model

[7] Yağanoğlu, M. (2022). Hepatitis C virus data analysis and prediction using machine learning. Data Knowledge Engineering, 142, 102087.

[8] P. Marcellin, M. Ziol, P. Bedossa, et al., "Non-invasive assessment of liver fibrosis by stiffness measurement in patients with chronic hepatitis B," Liver International, vol. 29, no. 2, pp. 242-247, 2009

[9] T. Poynard, R. Morra, P. Halfon, et al., "Meta-analyses of FibroTest diagnostic value in chronic liver disease," BMC Gastroenterology, vol. 7, no. 40, 2007.

[10] M. Ziol, A. Handra-Luca, A. Kettaneh, et al., "Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C," Hepatology, vol. 41, no. 1, pp. 48-54, 2005.

[11] C. Jiang, Z. Xu, J. Liu, R. Li, K. Chen, W. Peng, Y. Xiao, D. Cheng, L. Fu, S. Peng, "Noninvasive diagnosis of significant liver fibrosis in patients with chronic hepatitis B using nomogram and machine learning models," Scientific Reports, vol. 15, no. 1, p. 571, 2025.

[12] R. Anteby, E. Klang, N. Horesh, et al., "Deep learning for noninvasive liver fibrosis classification: A systematic review," Liver International, vol. 41, pp. 2269-2278, 2021.

[13] Y. Wang, L. Chen, Z. Fan, et al., "A radiomics-based deep learning model for non-invasive assessment of liver fibrosis in chronic hepatitis B patients," EBioMedicine, vol. 85, p. 104293, 2022.

[14] X. Zhang, T. Li, M. Chen, et al., "External validation and population-specific calibration of non-invasive fibrosis models in chronic hepatitis B: A multicenter cohort study," Hepatology International, vol. 18, no. 1, pp. 60-72, 2024.

[15] Y. Liu, J. Wang, L. Yang, et al., "Sequential combination of serum markers and transient elastography for the assessment of liver fibrosis in chronic hepatitis B patients: A prospective multicenter study," Liver International, vol. 43, no. 9, pp. 2176-2185, 2023.

[16] S.U. Kim, H.S. Kim, J.Y. Park, et al., "Machine learning-based model for prediction of significant fibrosis in chronic hepatitis B using serum biomarkers," Journal of Hepatology, vol. 78, no. 2, pp. 335-344, 2023.