Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2202~2211
e-ISSN: 2548-6861

Enhancing Liver Cirrhosis Staging Accuracy using Optuna-Optimized TabNet

Muhammad Farhan Arifin [1]*, Ika Novita Dewi [2]*, Abu Salam [3], Danang Wahyu Utomo [4]**, Sindhu Rakasiwi [5]****
* Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang
** Research Center for Intelligent Distributed Surveilance and Security (IDSS), Universitas Dian Nuswantoro, Semarang
112202206917@mhs.dinus.ac.id [1], ikadewi@dsn.dinus.ac.id [2], abu.salam@dsn.dinus.ac.id [3], danang.wu@dsn.dinus.ac.id [4], sindhu.rakasiwi@dsn.dinus.ac.id [5]

Article Info

Article history:

Received 2025-09-02
Revised 2025-09-10
Accepted 2025-09-19

Keyword:

Deep Learning,
Liver Cirrhosis,
Optuna,
TabNet,
Tabular Data.

ABSTRACT

Liver cirrhosis is a progressive chronic disease whose early detection poses a clinical challenge, making accurate severity staging crucial for patient management. This research proposes and evaluates a TabNet deep learning model, specifically designed for tabular data, to address this challenge. In the initial evaluation, a baseline TabNet model with its default configuration achieved a baseline accuracy of 65.11% on a public clinical dataset. To enhance performance, hyperparameter optimization using Optuna was implemented, which successfully increased the accuracy significantly to 70.37%, with precision, recall, and F1-score metrics each reaching 70%. The model's discriminative ability was also validated as reliable in multiclass classification through AUC metric evaluation. In addition to accuracy improvements, the model's interpretability was validated through the identification of key predictive features such as Prothrombin and Hepatomegaly, which align with clinical indicators. This study demonstrates that Optuna-optimized TabNet is an effective and interpretable approach, possessing significant potential for integration into clinical decision support systems to support a more precise diagnosis of liver cirrhosis.

I. INDROBUCTION

Liver cirrhosis is a chronic medical condition characterized by progressive, long-term damage to liver tissue, wherein healthy tissue is gradually replaced by excessive fibrosis. This process disrupts the liver's normal structure and function, potentially leading to liver failure and hepatocellular carcinoma [1]. As a leading cause of mortality globally, particularly in developing nations, its prevalence is driven by risk factors such as chronic alcohol consumption, viral hepatitis, and Non-Alcoholic Fatty Liver Disease (NAFLD) [2]. The significance of liver cirrhosis is amplified by its often-asymptomatic nature in the early stages, with detection frequently occurring only at an advanced stage. This delay complicates medical intervention, reduces treatment efficacy, and imposes a substantial burden on healthcare systems [3].

According to the World Health Organization (WHO), liver cirrhosis affects approximately 3% of the global population, with an estimated 3 to 4 million new cases diagnosed annually. The etiology of the disease exhibits significant regional variation; in Indonesia, hepatitis B and C are the predominant causes, whereas in Western countries, chronic alcohol consumption is the primary driver [4]. This global health issue underscores the critical need for advanced and precise scientific approaches for early detection and prevention to mitigate mortality.

Challenges in managing liver cirrhosis are exacerbated by delayed diagnosis, often stemming from a lack of regular public health screenings [5]. Suboptimal management of the disease leads to high mortality rates and places a significant strain on healthcare resources. While early detection is crucial, it remains difficult without accurate predictive tools [6]. Neglected cases of cirrhosis can progress to fatal liver cancer, with elderly individuals and patients with chronic hepatitis being the most vulnerable populations [7].

Previous studies have attempted to use machine learning approaches to predict the risk of liver cirrhosis. For example, a study by Yasmin Roni Mz (2024) used the CART method to generate a Decision Tree algorithm on a liver cirrhosis

prediction dataset, resulting in an accuracy of less than 70% in identifying patients with potential cirrhosis [8]. A subsequent study by Vania Riskasari YR (2023) used the support vector machine (SVM) algorithm and demonstrated good performance with an accuracy of 67.68% [9]. Meanwhile, the latest study conducted by Mardewi (2024) using the Ensemble Bagged Tree model achieved the highest accuracy of 71%, followed by Ensemble Boosted Tree (67.2%) and Ensemble RUSBoosted Tree (66%) for classifying patients with various stages of liver cirrhosis [10].

Although various approaches have been used for the prediction and classification of liver cirrhosis, there are still a number of research gaps that need to be addressed. First, research by Yasmin Roni Mz (2024) using the CART method resulted in low accuracy (<70%) due to data imbalance and dependence on data structure, without cross-validation or comparison with other models. Second, Vania Riskasari YR (2023) used SVM with an accuracy of 67.86%, but did not optimally handle minority classes and only relied on the RBF kernel without exploring data balancing techniques. Third, Mardewi and Supriyadi La Wungo (2024) compared several ensemble models, but their evaluation was still limited to accuracy, without considering metrics such as F1-score or AUC, and did not apply an Explainable AI (XAI) approach. Additionally, there has been no research on long term cirrhosis progression prediction or survival analysis. In medical research for classification, machine learning methods have been widely used as a general approach to create accurate classifications and predictions, but their performance depends on the dataset used [11]. On the other hand, although most studies have used real clinical data, none have specifically explored the potential of the TabNet deep learning model designed for tabular data, which offers advantages in feature selection and interpretability. Therefore, further research is needed to explore more adaptive models and the use of architectures such as TabNet to improve the accuracy and applicability of liver cirrhosis predictions.

In response to this architectural gap, TabNet has emerged as a state of the art model specifically engineered for tabular data. It uniquely integrates concepts from decision trees into a deep learning framework, employing a sequential attention mechanism to perform instance-wise feature selection. This design not only enables TabNet to capture complex patterns but also provides a degree of interpretability, a crucial feature for clinical applications. Given these advantages, TabNet presents a compelling candidate for addressing the challenges of liver cirrhosis classification. This research aims to fully leverage this potential by implementing a deep learning based framework for liver cirrhosis prediction, utilizing the TabNet model with hyperparameter optimization. TabNet is specifically engineered for tabular data, demonstrating robust capabilities in capturing complex patterns and relationships within structured clinical patient data [12]. TabNet's inherent ability to perform automatic feature selection can reduce manual workload and increase implementation efficiency [13]. However, even a powerful architecture like TabNet is not guaranteed to perform optimally out of the box, as its performance is highly sensitive to the choice of hyperparameters. To systematically navigate this complex parameter space and unlock TabNet's full predictive potential, this research employs Optuna, a state of the art Bayesian optimization framework. The integration of Optuna is critical for achieving optimal model accuracy and resilience through dynamic parameter searching [14].

The primary objective of this research is to develop and systematically evaluate an Optuna optimized TabNet model for the multiclass classification of liver cirrhosis severity from clinical tabular data. The novelty of this research lies in its specific application and rigorous optimization of the TabNet architecture, a model engineered for tabular data, for this critical clinical challenge, an area that remains underexplored in the existing literature. By doing so, this study contributes a robust, high performing, and interpretable framework that not only outperforms a baseline model but also provides a validated pathway for enhancing diagnostic precision in liver cirrhosis management.

II. METHOD

This research aims to classify the severity of liver cirrhosis in target stages that indicate the disease has reached stage one, two, or three by developing and evaluating the TabNet model, which is part of a deep learning classification model [15]. The research methodology, as illustrated in figure 1, encompasses a multi-stage workflow. This process begins with dataset preparation and preprocessing, followed by data splitting. Subsequently, the TabNet model is trained and then refined through hyperparameter optimization using Optuna, with a final evaluation conducted on both the training and testing sets.

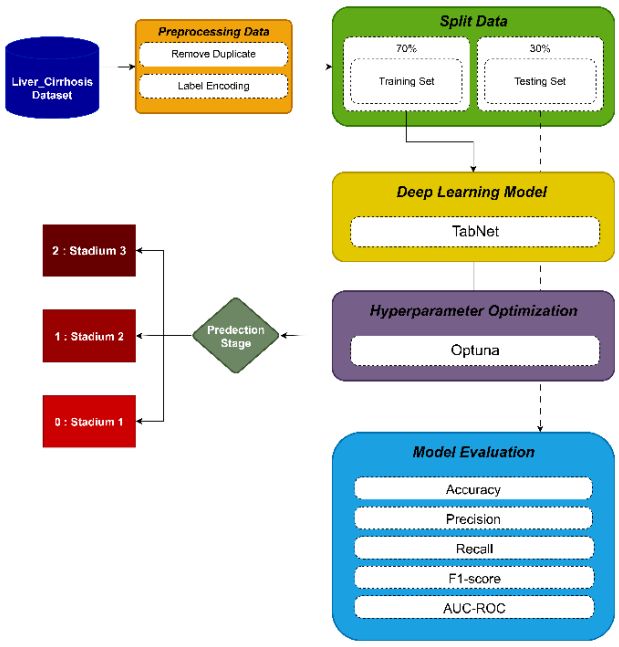

Figure 1. Research flow

A. Dataset Preparation

This research used a dataset from Kaggle (https://www.kaggle.com/datasets/aadarshvelu/liver-cirrhosis-stage-classification) that contains medical information about patients with liver cirrhosis. It consists of 25,000 rows of data with 19 features, and 3 class labels of stage (1,2,3) including numeric data (N_Days, Age, Bilirubin, Cholesterol, Albumin, Copper, Alk_Phos, SGOT, Triglycerides, Platelets, Prothrombin, and Stage) and non-numeric data (Status, Drug, Sex, Ascites, Hepatomegaly, Spiders, and Edema).

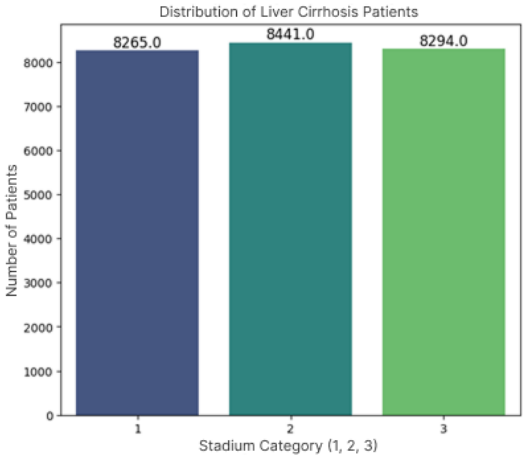

Figure 2 . Dataset distribution

The distribution of classes within the dataset is balanced, as illustrated in figure 2, which provides a solid foundation for building classification models without significant class imbalance bias. However, despite the relatively large initial size, the effective dataset became limited after preprocessing steps, particularly the removal of 15,361 duplicate entries, leaving only 9,639 unique data points available for model training and evaluation. This considerable reduction in dataset size imposes limitations on the model's capacity to generalize and fully capture the clinical heterogeneity present in liver cirrhosis patients. Although class balance is maintained within the reduced data, the smaller volume increases the vulnerability to overfitting and may reduce the robustness of the model when applied to unseen clinical data. Consequently, this highlights the necessity for future research to incorporate larger, more diverse, and representative clinical datasets to enhance model reliability and applicability in real-world medical settings.

B. Data Preprocessing

Preprocessing data is a crucial initial stage in building a deep learning model because good data quality greatly affects the model's performance. In this research, the preprocessing stage consisted of two steps:

1) *Remove Duplicate*: This step aims to remove duplicate data rows in the dataset. This dataset has 15,361 duplicate data points. The remaining data used is 9,639 data points.

2) *Label Encoding*: Label encoding is a technique for converting categorical features in the form of text into numerical values [16]. Features that will be encoded include non-numeric features consisting of Status, Drug, Sex, Ascites, Hepatomegaly, Spiders, and Edema.

C. Split Data

The preprocessed dataset was partitioned into a 70% training set and a 30% testing set using a stratified split to maintain the original class distribution in both subsets. This approach ensures an unbiased evaluation of the model's ability to generalize to unseen data [17].

Adhering to a 70/30 split ratio, the dataset was divided into two distinct subsets: a training set and a testing set. The training set, comprising 6,747 instances, was utilized exclusively for the model development and learning phase. The remaining 2,892 instances constituted the testing set, which was held out to provide an unbiased assessment of the final model's performance.

D. TabNet Model

TabNet is a deep learning framework designed for tabular data, integrating representation learning with interpretability. Through sequential attention and sparse feature selection, it adaptively identifies the most informative features at each decision step, enhancing both predictive accuracy and efficiency [18].

The main advantage of TabNet is its ability to maintain transparency in the decision-making process through feature attribution visualization mechanisms [19]. Additionally, TabNet is suitable for application in the medical domain, as it addresses the interpretability limitations of conventional deep learning and provides accurate predictive results on heterogeneous data such as patient data [16].

As shown in figure 3, at each step of the decision making process, TabNet employs a specific formula to determine the most significant features, as outlined figure 3.

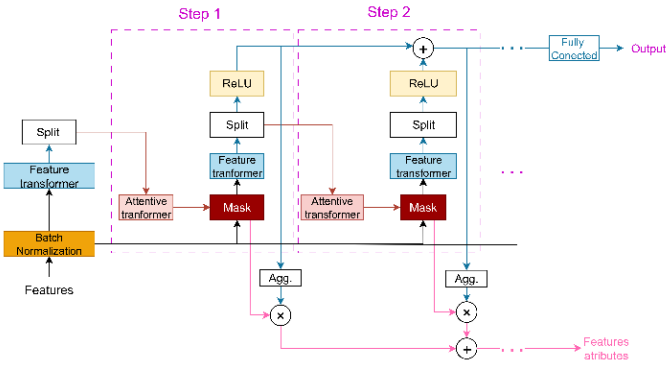

Figure 3. TabNet architecture framework

First, the data passes through a batch normalization layer before being processed by the feature transformer. Extracting

features from the normalized input is the responsibility of the feature transformer module. Each feature transformer is divided into two parts: the first part uses shared parameters, while the second part has independent parameters and is trained specifically for each step [15]. This architecture consists of two parts, each of which has a fully connected layer followed by batch normalization (BN) and gated linear units (GLU). The output from each part is normalized again to maintain variance stability and facilitate learning within the TabNet architecture. The TabNet architecture is composed of three main components. The Feature Transformer processes features into initial latent representations, which are then passed to the Attentive Transformer. The Attentive Transformer identifies the most relevant features to focus on in the subsequent decision step. Equation (1) is the formula for the TabNet mechanism.

$$M[i] = sparsemax(P[i-1] \cdot h_i(a[i-1])) \quad (1)$$

$a[i-1]$ is the characteristic information shared by the split in the final decision, h_i describes the BN and FC layers, $P[i-1]$ is the feature usage priority scale, and sparsemex can produce less frequent output results [20]. Thus, TabNet can provide good performance by performing sparse feature selection to support model interpretability [21]. In the context of liver cirrhosis classification, the application of TabNet is relevant because patient clinical data is generally multidimensional tabular data with complex inter-feature relationships, and requires a system that is not only accurate but also clinically accountable.

E. Hyperparameter Optimization

The optimization of a model's hyperparameters is essential [22], as these factors critically determine its learning ability and generalization performance. In this research, Optuna Model was utilized for optimalization parameter. Built around the "define-by-run" principle, Optuna supports a dynamic and flexible approach to parameter exploration. The framework incorporates efficient search mechanisms and pruning strategies, along with a highly adaptable architecture suitable for diverse optimization tasks.

The methodological approach consisted of three stages. Initially, the hyperparameters and their specific search spaces were meticulously defined for each model. Subsequently, an objective function was formulated to maximize model performance, and the total number of optimization trials (*n_trials*) was specified. The final stage involved a thorough evaluation of the optimization outcomes to identify the optimal hyperparameter configurations.

F. Evaluation Model

This research used several key classification metrics to evaluate the data: accuracy, precision, recall, F1-score, and area under the curve (AUC-ROC). Equation (2) is the accuracy formula for the model performance evaluation metric.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision measures how accurate the model is when predicting positive cases, whereas recall shows how well the model identifies all actual positive instances. The formulas for calculating both precision and recall are provided in Equation (3) as part of the performance evaluation metrics.

$$precision = \frac{TP}{TP+FP}, recall = \frac{TP}{TP+FN} \quad (3)$$

The F1-score is the harmonic mean of precision and recall, which provides a more balanced assessment of the two metrics. Equation (4) is the F1-Score formula for model performance evaluation metrics.

$$F1 - Score = 2 \times \frac{Precesion \times Recall}{Precision + Recall} \quad (4)$$

The ROC curve is a graphical tool that illustrates the relationship between sensitivity (TPR) and false positive rate (FPR) under different classification thresholds [26]. Sensitivity indicates the proportion of positives correctly detected, while FPR shows negatives incorrectly classified as positives. This value can be calculated using the equation shown in Formula (5).

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (5)$$

III. RESULT AND DISCUSSION

This section details the implementation and evaluation of an Optuna-optimized TabNet model for classifying liver cirrhosis severity.

A. Result of TabNet Baseline Model

The initial phase of the study focused on establishing a baseline performance for the TabNet model. While the core TabNet architecture was utilized with its default settings, the training process was governed by a specific set of parameters, as detailed in table I. The model was trained for a maximum of 50 epochs, with an early stopping patience of 10 epochs implemented to prevent overfitting. A batch size of 256 and a virtual batch size of 128 were used to balance computational efficiency and gradient stability. This configuration was intentionally chosen to provide a standardized, out-of-the-box performance benchmark before proceeding with hyperparameter optimization.

TABLE I
PARAMETERS SETTINGS

TabNet Model	Parameters
max_epochs	50
patience	10
batch_size	256
virtual_batch_size	128

The performance of the baseline TabNet model is detailed in table II. Table II presents the classification performance of the baseline TabNet model on the testing set. The model achieved an overall accuracy of 65.11%. The macro and weighted averages for precision, recall, and F1-score were highly consistent at approximately 0.65-0.66, indicating that the model's performance was relatively balanced across the three classes, as expected given the similar support values for each class. A closer examination of the per-class metrics reveals varying performance: the model was most effective at classifying Class 2 (F1-score: 0.72) but struggled significantly with Class 1, which had a notably low precision of 0.57. Despite the moderate accuracy, the Area Under the Curve (AUC) of 0.8263 suggests the model possesses good discriminative power. Overall, these results establish a reasonable but clearly improvable performance baseline, with specific weaknesses to be addressed through hyperparameter optimization.

TABLE II
CLASSIFICATION PERFORMANCE OF THE BASELINE MODEL

	Precision	Recall	F1-score	Support
0	0.65	0.62	0.63	908
1	0.57	0.65	0.61	992
2	0.75	0.69	0.72	992
macro avg	0.66	0.65	0.65	2892
weighted avg	0.66	0.65	0.65	2892
AUC	0.8263			
Accuracy	0.6511			

The confusion matrix for the baseline model, shown in figure 4, illustrates its performance in the multiclass classification task. The main diagonal highlights a substantial number of correct predictions, with the model accurately classifying 559, 640, and 684 samples for Class 0, 1, and 2, respectively. However, the off-diagonal values reveal a significant systematic bias. The model demonstrates a clear propensity to misclassify samples into Class 1 (moderate stage), incorrectly assigning 270 samples from Class 0 and 211 samples from Class 2 to this category. This pattern suggests that the feature distributions of Class 1 may overlap significantly with those of the adjacent classes, making it a common point of confusion for the baseline model. While the number of correct classifications confirms the model's fundamental effectiveness, this pronounced bias highlights a critical area for improvement through hyperparameter optimization.

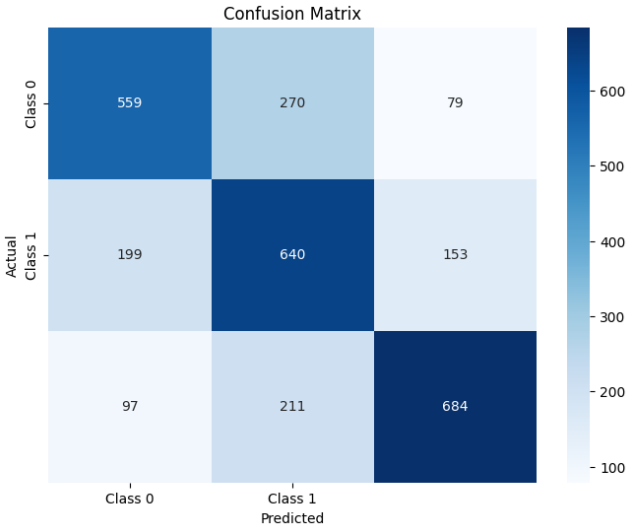

Figure 4. Confusion matrix of the baseline model

The model's per-class discriminative performance was evaluated using One-vs-Rest (OvR) ROC curves, as shown in figure 5. The model demonstrates strong performance in distinguishing Class 2 (severe) from the others, achieving the highest AUC of 0.87, and also performs well on Class 0 (mild) with an AUC of 0.84. However, the model's ability to discriminate Class 1 (moderate) is notably weaker, with an AUC of 0.77. This lower score for the intermediate class aligns with the findings from the confusion matrix (Figure 4), which revealed a significant tendency for the model to misclassify samples from both Class 0 and Class 2 into this category. Nevertheless, as all curves lie well above the diagonal line of random chance, the results confirm that the baseline model possesses significant predictive power overall.

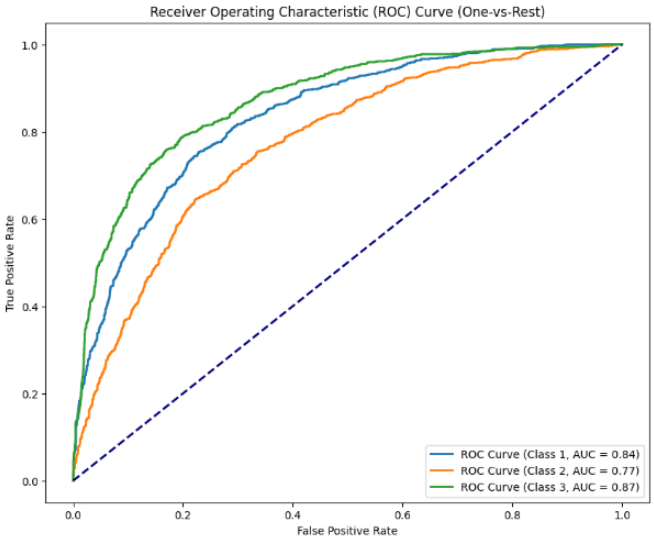

Figure 5. ROC curve of the baseline model

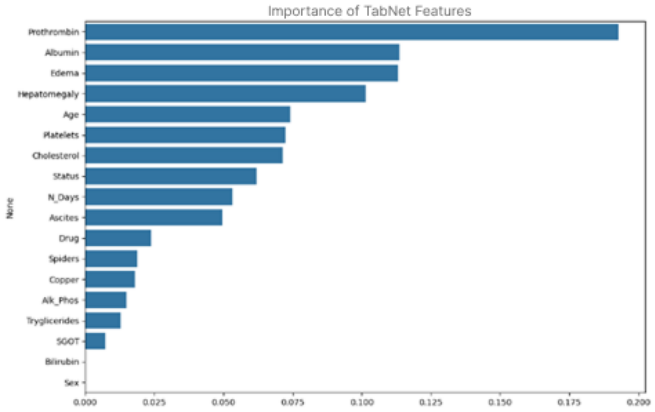

Figure 6. Important features in the Tabnet baseline model

The feature importance analysis for the baseline TabNet model, illustrated in figure 6, identifies a clear hierarchy of predictive features. The analysis reveals a clear hierarchy of feature influence, which can be categorized into three distinct tiers. The top tier of predictors, with importance scores approaching 0.20, is dominated by clinically critical markers of liver function and decompensation: 'Prothrombin', 'Albumin', and 'Edema'. The high ranking of these features aligns with established medical knowledge, as they directly reflect the liver's synthetic capacity and the presence of fluid retention, key indicators of cirrhosis severity. An intermediate tier of features, including 'Hepatomegaly', 'Age', 'Platelets', and 'Cholesterol', also demonstrates significant predictive value. These features represent a mix of physical examination findings, demographic data, and secondary laboratory markers that provide important contextual information for staging. Finally, a large number of features, such as 'SGOT', 'Bilirubin', and 'Sex', exhibited minimal to negligible importance, indicating that the TabNet model's attention mechanism effectively learned to de-prioritize less relevant or redundant information. This instance-wise feature selection is a key strength of the TabNet architecture.

B. Model Optimization Results with Optuna

While the baseline TabNet model demonstrated a foundational predictive capability, its initial performance was deemed unsatisfactory, highlighting the need for systematic optimization. Therefore, to enhance the model's accuracy, stability, and generalization, hyperparameter optimization was conducted using the Optuna framework. This subsequent phase aimed to identify the optimal set of parameters to unlock the full potential of the TabNet architecture on this specific dataset.

TABLE IV
OPTUNA PARAMETERS SETTINGS

Parameters	Range
n_d	138
n_a	146
n_steps	4
gamma	1.169
lambda_sparse	6.126e-06
momentum	0.104
clip_value	2.346
learning rate	0.037
weight_decay	5.337e-06
batch_size	128

Following the hyperparameter optimization process with Optuna, the optimal configuration for the TabNet model was identified, as detailed in table IV. The optimal architecture was defined by an attention dimension (n_a) of 146 and a feature dimension (n_d) of 138, with 4 decision steps (n_steps). Key training parameters were fine-tuned to balance learning and regularization: the learning rate was set to 0.037, with a momentum of 0.104 and a gamma of 1.169. To prevent overfitting, regularization parameters were optimized to lambda_sparse = 6.126e-06 and weight_decay = 5.337e-06, while a clip_value of 2.346 ensured gradient stability. Finally, a batch size of 128 was found to be optimal. This specific combination of parameters, discovered through Optuna's efficient exploration, represents the configuration that yielded the highest model performance.

The accuracy evaluation of the optimized model is presented in figure 7. The graph shows a rapid and parallel increase in both accuracies during the initial epochs, indicating efficient learning from the data. As training progresses, the curves begin to plateau, with the training accuracy converging to approximately 0.74-0.75 and the validation accuracy stabilizing around 0.69-0.70. The close proximity and parallel trajectory of the two curves are particularly noteworthy, as they demonstrate strong generalization and confirm that the model is not overfitting to the training data. This stability in the later epochs indicates that the model has reached convergence, and that the optimized parameters have successfully guided the learning process to a robust solution.

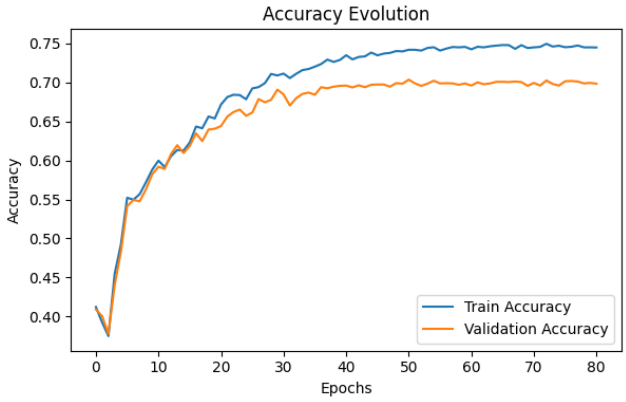

Figure 7. Optuna-optimized TabNet model accuracy

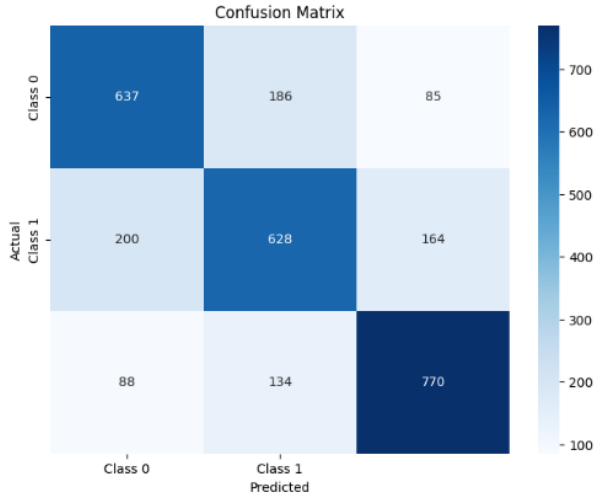

Figure 8. Confusion matrix of the optuna-optimized TabNet model

The confusion matrix for the optimized model, depicted in figure 8, demonstrates a significant enhancement in classification accuracy and reliability. The main diagonal shows a marked increase in true positives, with the model now correctly classifying 637, 628, and 770 samples for Class 0, 1, and 2, respectively. A key improvement is the significant reduction in the systematic bias towards Class 1; for instance, the misclassification of Class 0 as Class 1 dropped from 270 to 186 instances. While some confusion between adjacent classes persists, the overall number of misclassifications has decreased considerably. This enhanced performance, characterized by a significant reduction in false positives and false negatives, confirms that the optimized model is not only more accurate but also more balanced and reliable, making it a stronger candidate for providing trustworthy decision support in clinical settings.

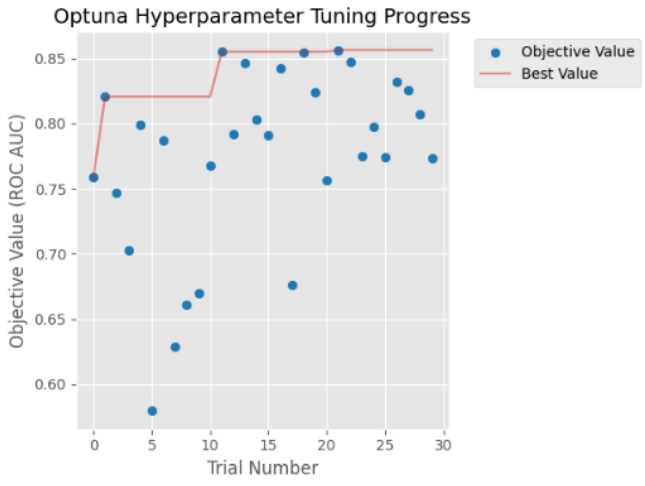

Figure 9. Optuna-optimized TabNet model hyperparameter

The hyperparameter optimization process, facilitated by the Optuna framework, is illustrated in figure 9. The graph clearly demonstrates the effectiveness of the Tree-structured Parzen Estimator (TPE) algorithm, which intelligently guides the search toward more promising regions of the parameter space. This is evidenced by the upward trend in the best-achieved score (the red line), which converges as the number of trials increases. The concentration of high-scoring trials in later stages indicates that Optuna successfully learned from previous results to focus on high-performance configurations. Furthermore, this guided search, combined with Optuna's ability to prune unpromising trials early, resulted in a more computationally efficient optimization process compared to traditional methods like grid or random search, ultimately leading to a superior final model.

The feature importance analysis of the optimized TabNet model, illustrated in figure 10, identifies a distinct hierarchy of predictive features. The analysis reveals a clear hierarchy of feature influence, with 'Prothrombin' emerging as the single most dominant predictor, having an importance value approaching 0.20. Following this, a primary tier of highly significant features includes 'Hepatomegaly' and 'Albumin', both of which are established clinical indicators of liver function and disease progression.

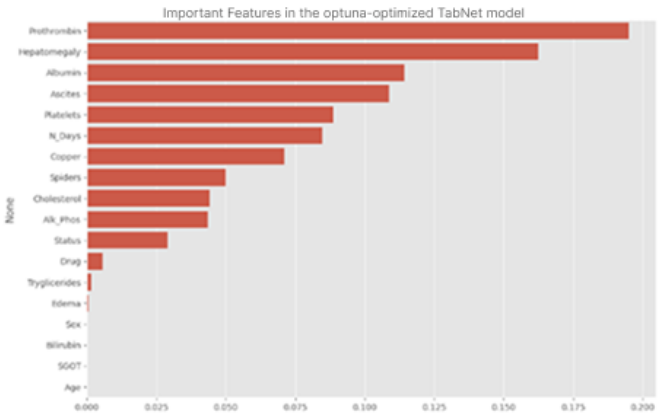

Figure 10. Important features in the optuna-optimized TabNet model

A secondary tier of moderately impactful features was also identified, including 'Ascites', 'Platelets', 'N_Days', and 'Copper', with importance values ranging from approximately 0.07 to 0.12. The contribution of these features underscores the model's ability to integrate a combination of physical symptoms and laboratory results. In contrast, a large number of other clinical and demographic variables, such as 'SGOT', 'Bilirubin', and 'Age', were assigned minimal to negligible importance. This demonstrates the model's capacity for instance-wise feature selection, effectively de-prioritizing less relevant information. The resulting feature hierarchy not only confirms the model's alignment with clinical knowledge but also provides a valuable framework for developing data-driven medical decision support systems.

C. Performance Comparison: Baseline TabNet vs. Optuna-Optimized TabNet

This section presents a direct comparative analysis of the TabNet model's performance before and after hyperparameter

optimization with Optuna. To quantify the impact of the optimization process, the performance of the baseline model is benchmarked against the final, Optuna-optimized model across all primary evaluation metrics. As the results in table V demonstrate, the optimization led to a significant and consistent improvement in accuracy, precision, recall, and F1-score.

TABLE V
BASELINE TABNET VS. OPTUNA-OPTIMIZED TABNET

Model	Accuracy	Precision	Recall	F1-Score
TabNet	65,11%	66%	65%	65%
TabNet + Optna	70,37%	70%	70%	70%

As summarized in table V, the hyperparameter optimization with Optuna yielded a significant improvement in the TabNet model's performance. The baseline model established a performance benchmark with an accuracy of 65.11% and an F1-score of 65%. Following optimization, the model's performance improved substantially across all metrics, with accuracy increasing by 5.26% to 70.37%. This substantial gain is attributed to the transition from generic, default hyperparameters to a set specifically tailored to the unique characteristics of the dataset. Optuna's intelligent search algorithm navigated TabNet's complex parameter space to find a configuration that strikes a superior bias-variance trade-off, enabling the model to better capture the underlying data patterns without overfitting. This targeted optimization is what unlocked the model's true predictive potential, leading directly to the enhanced accuracy and F1-score.

Notably, the optimized model achieved a balanced performance, with precision, recall, and F1-score all converging at 70%. This parity between precision and recall is a critical outcome, indicating that the model is equally adept at correctly identifying positive cases and minimizing false positives. The consistent F1-score of 0.70 further underscores this harmonization of metrics, confirming that the Optuna-optimized configuration produced a more robust and reliable classifier, which is essential for clinical decision-making applications.

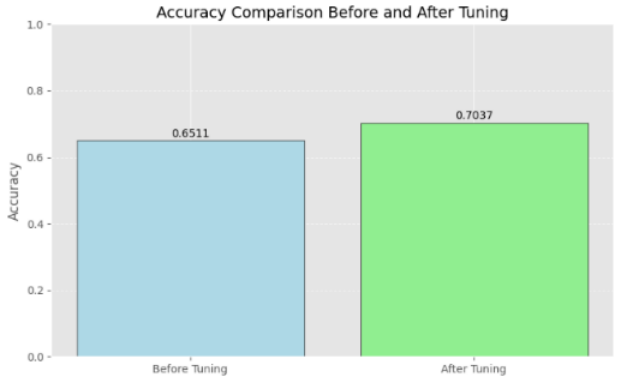

Figure 11. TabNet model comparison accuracy before and after optimization

Figure 11 provides a visual summary of the substantial performance improvement achieved through hyperparameter tuning. The baseline model achieved an accuracy of 65.11%, which rose to 70.37% following optimization with Optuna. This represents an absolute improvement of 5.26 percentage points, directly demonstrating the efficacy of Optuna in navigating the complex parameter space to refine the model's behavior.

This accuracy improvement is not merely a numerical gain; it signifies enhanced generalization and increased model stability. The optimized model is better equipped to handle the inherent complexity and variations within the liver cirrhosis dataset. In a clinical context, such an improvement, while seemingly marginal, can be highly significant, as it translates to a more reliable diagnostic tool that can support more confident clinical decision-making.

This study not only highlights the effectiveness and interpretability of TabNet-Optuna but also serves as a pioneering benchmark for using tabular deep learning models in liver cirrhosis staging. This advancement creates significant opportunities for integrating artificial intelligence (AI) into clinical decision support systems (CDSS). By incorporating this model, physicians can receive more objective and consistent recommendations for diagnosis and staging, aiding in precise treatment planning and better risk assessment. Such integration can reduce human errors in interpreting lab results and clinical signs, which are often subjective, thereby enhancing the quality of medical decisions. Ultimately, this improves diagnostic reliability, streamlines patient care, and supports the broader adoption of AI as a tool for augmented intelligence in daily clinical practice.

D. Error Analysis

This analysis included the identification of common patterns in classification errors, particularly in cases that were repeatedly misclassified between liver cirrhosis stage classes. The confusion matrix results before optimization showed that the model tended to be biased toward class 1 (moderate stage), with many cases from class 0 (mild stage) and class 2 (severe stage) being misclassified as class 1. This indicates an overlap of clinical features between the mild and severe stages and the moderate stage, resulting in less decisive model decision making. Additionally, the relatively lower AUC performance in class 2 before optimization (0.77) reinforces the suspicion that the model struggles to distinguish the characteristic features of patients with advanced cirrhosis.

After optimization with Optuna, there was an improvement in classification accuracy and balance across all three classes. However, misclassification still occurred, particularly in the transition between class 0 and class 1. Analysis of misclassified instances revealed that most of these data points had overlapping feature threshold values, such as albumin and platelet levels that fell within the normal range but also appeared in patients with mild and moderate stages. This indicates that ambiguity in clinical features with

overlapping value distributions remains a significant challenge.

Additionally, although the "Prothrombin" and "Hepatomegaly" features were consistently identified as the most important features by TabNet, there were specific cases where the values of these features were within the normal range but the data was classified into the wrong stage. This phenomenon suggests that over-reliance on dominant features without considering the combination of other contextual features can reduce prediction accuracy. In summary, the error analysis highlights the need to improve data balancing strategies, outlier detection, and the integration of explainable AI methods to increase the model's accuracy and reliability in a clinical setting.

E. Comparision with Previous Research

To contextualize the performance of the Optuna-optimized TabNet model, a comparison was made with results from previous studies on similar liver cirrhosis classification tasks, as summarized in table VI. The optimized TabNet model achieved an accuracy of 70.37%, positioning it competitively against established conventional and ensemble methods. Specifically, the model comfortably outperforms the Support Vector Machine approach (67.68%) and achieves performance on par with the CART method (>70%).

Notably, its accuracy is nearly equivalent to that of the more complex Ensemble Bagged Tree model (71%). This is a significant finding, as it demonstrates that a single, interpretable deep learning architecture, when properly optimized, can match the predictive power of ensemble methods, which often sacrifice interpretability for performance. This competitive performance is attributed to TabNet's inherent capacity to capture complex, non-linear relationships within the data, a capability that was fully unlocked through systematic hyperparameter optimization.

TABLE VI
PERFORMANCE COMPARISON WITH PREVIOUS RESEARCH

Reference	Method	Accuracy
[8]	CART Decision Tree	>70%
[9]	Support Vector Machine	67,68%
[10]	Ensamble Bagged Tree	71%
Ours	TabNet + Optuna	70,37%

F. Limitation of the Study

This study successfully showed that using TabNet optimized with Optuna improved accuracy and interpretability, but there are several limitations to consider. Firstly, the model was only tested on publicly available secondary datasets and has not been validated with actual clinical data from medical institutions. This restricts the model's ability to generalize, as public data may not fully capture the wide range of patient characteristics found in real populations.

Secondly, the study did not perform external validation using independent datasets from different healthcare facilities. External validation is crucial to confirm the model's robustness, reliability, and readiness for clinical application. Without it, the model may still suffer from biases due to limited data representation, potentially reducing prediction accuracy in real-world settings.

Therefore, future research should aim to use more varied local clinical data and carry out external validations across multiple centers. This will provide stronger proof that the TabNet-Optuna model not only excels in experiments based on public datasets but also has practical significance and real clinical value in supporting AI-driven medical decisions.

IV. CONCLUSION

This research aims to classify the severity of liver cirrhosis based on tabular clinical data using the TabNet deep learning model optimized with Optuna. The initial TabNet model showed an accuracy of 65.11% with relatively balanced performance across the three classification classes. After hyperparameter optimization using Optuna, the model's performance improved significantly, achieving an accuracy of 70.37% and improvements in all primary evaluation metrics, including precision, recall, and F1-score, each reaching 70%. Additionally, the TabNet model provides good interpretability through feature importance mapping, where features such as Prothrombin, Hepatomegaly, and Albumin were identified as the most significant indicators in the diagnosis of liver cirrhosis.

However, this research has limitations in terms of the scope of testing, which is still limited to secondary datasets and does not involve direct clinical data from local medical institutions. In addition, although hyperparameter optimization was performed using Optuna, the approach used still focuses on a single model without direct comparison with other deep learning architectures. Therefore, future development efforts could focus on utilizing larger and more representative clinical datasets, exploring data balancing methods, and applying alternative architectures such as Transformer-based tabular models to improve prediction accuracy and model robustness against medical data variations. Overall, the findings from this research indicate that TabNet not only excels in accuracy performance after optimization but also offers advantages in terms of interpretability and clinical relevance through feature selection. Thus, the optimized TabNet model has significant potential for implementation in clinical decision support systems, particularly in supporting more efficient, accurate, and data-driven early detection and risk management of liver cirrhosis.

REFERENCES

[1] P. Ginès, A. Krag, J. G. Abraldes, E. Solà, N. Fabrellas, and P. S. Kamath, "Liver cirrhosis," Oct. 2021, *Elsevier B.V.* doi: 10.1016/S0140-6736(21)01374-X.

[2] R. Davis, "Examining the immune response to infection in the context of Non-alcoholic fatty liver disease (NAFLD),"

https://ucalgary.scholaris.ca/items/44ad3654-ae2e-4123-9cc4-66a0cdb5d2fe.

[3] Z. M. Younossi, G. Wong, Q. M. Anstee, and L. Henry, "The Global Burden of Liver Disease," Jul. 2023, *W.B. Saunders*. doi: 10.1016/j.cgh.2023.04.015.

[4] B. P. Silaban, F. Lumongga, and H. Silitonga, "Karakteristik Penderita Sirosis Hati," https://ejurnal.methodist.ac.id/index.php/jkm/article/view/1322.

[5] E. B. Tapper and N. D. Parikh, "Diagnosis and Management of Cirrhosis and Its Complications: A Review," May 2023, *American Medical Association*. doi: 10.1001/jama.2023.5997.

[6] M. W. Sonderup *et al.*, "Managing cirrhosis with limited resources: perspectives from sub-Saharan Africa," Feb. 2024, *Elsevier Ltd*. doi: 10.1016/S2468-1253(23)00279-0.

[7] C. K. Kang, P. N. Brennan, and J. F. Dillon, "How to Effectively Monitor Aging Patients with Chronic Hepatitis B: A Review," 2022, *Dove Medical Press Ltd*. doi: 10.2147/CIA.S366255.

[8] Yasmin Roni Mz, I Komang Gde Sukarsa, and I Gusti Ayu Made Srinadi, "Klasifikasi Penyakit Sirosis Hati Dengan CART," *Jurnal Cahaya Mandalika ISSN 2721-4796 (online)*, vol. 5, pp. 308–315, Mar. 2024, doi: 10.36312/jcm.v5i2.2123.

[9] V. Riskasari YR, I. P. E. N. Kencana, and I. K. G. Sukarsa, "Klasifikasi Penyakit Sirosis Menggunakan Support Vector Machine," *E-Jurnal Matematika*, vol. 12, no. 2, p. 87, May 2023, doi: 10.24843/mtk.2023.v12.i02.p404.

[10] M. Mardewi and S. La Wungo, "Klasifikasi Liver Cirrhosis Menggunakan Teknik Ensemble: Studi Perbandingan Model Boosted Tree, Bagged Tree, dan Rusboosted Tree," *Journal of System and Computer Engineering (JSCE)*, vol. 5, pp. 219–225, Jul. 2024, doi: 10.61628/jsce.v5i2.1302.

[11] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *Jurnal Masyarakat Informatika*, vol. 13, no. 1, pp. 33–44, 2022, doi: 10.14710/jmasif.13.1.42912.

[12] I. Shaik, P. V. Nagamani, Y. Manmode, S. Yadav, V. Degala, and G. S. Rao, "Mapping the dynamics of global sea surface nitrate using ocean color data," Dec. 2024, *Elsevier Ltd*. doi: 10.1016/j.dynatmoce.2024.101507.

[13] M. A. Setitra and M. Fan, "Detection of DDoS attacks in SDN-based VANET using optimized TabNet," *Comput Stand Interfaces*, vol. 90, Aug. 2024, doi: 10.1016/j.csi.2024.103845.

[14] P. S. Pravin, J. Z. M. Tan, K. S. Yap, and Z. Wu, "Hyperparameter optimization strategies for machine learning-based stochastic energy efficient scheduling in cyber-physical production systems," *Digital Chemical Engineering*, vol. 4, Sep. 2022, doi: 10.1016/j.dche.2022.100047.

[15] S. Arık and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," in *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, Association for the Advancement of Artificial Intelligence, 2021, pp. 6679–6687. doi: 10.1609/aaai.v35i8.16826.

[16] T. Kurniawan, L. Hermawanti, and A. N. Safriandono, "Interpretable Machine Learning with SHAP and XGBoost for Lung Cancer Prediction Insights," *Journal of Applied Informatics and Computing*, vol. 8, no. 2 SE-Articles, pp. 296–303, Nov. 2024, doi: 10.30871/jaic.v8i2.8395.

[17] Y. Irawan, R. Pramitasari, W. M. Ashari, and A. N. H. Yansyah, "Support Vector Machine Classification Algorithm for Detecting DDoS Attacks on Network Traffic," *Journal of Applied Informatics and Computing*, vol. 9, no. 4 SE-Articles, pp. 1945–1954, Aug. 2025, doi: 10.30871/jaic.v9i4.10003.

[18] W. Li, "TabNet for high-dimensional tabular data: advancing interpretability and performance with feature fusion," *IET Conference Proceedings*, vol. 2025, pp. 168–173, Apr. 2025, doi: 10.1049/icp.2025.1030.

[19] M. Maftoun, N. Shadkam, S. Somayeh, S. Komamardakhi, Z. Mansor, and J. Hassannataj Joloudari, "Malicious URL Detection using optimized Hist Gradient Boosting Classifier based on grid search method," 2024.

[20] T. Zhang, M. Da Huo, Z. Ma, J. Hu, Q. Liang, and H. Chen, "Prediction model of stock return on investment based on hybrid DNN and TabNet model," *PeerJ Comput Sci*, vol. 10, pp. 1–23, 2024, doi: 10.7717/PEERJ-CS.2057.

[21] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, and G. Castignani, "Deep learning in insurance: Accuracy and model interpretability using TabNet," *Expert Syst Appl*, vol. 217, May 2023, doi: 10.1016/j.eswa.2023.119543.

[22] T. M. Arkan, A. Sugiharto, and H. A. Wibawa, "Optimizing VGG16 Architecture with Bayesian Hyperparameter Tuning for Tomato Leaf Disease Classification," *Jurnal Masyarakat Informatika*, vol. 16, no. 2, pp. 162–173, 2025, doi: 10.14710/jmasif.16.2.73168.