

**ANALISIS OPTIMASI HYPERPARAMETER DAN FEATURE
SELECTION PADA MODEL XGBOOST UNTUK PREDIKSI
INFEKSI HEPATITIS C**

LAPORAN SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

NADIA MARTHA LEFI

22.11.5140

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS OPTIMASI HYPERPARAMETER DAN FEATURE
SELECTION PADA MODEL XGBOOST UNTUK PREDIKSI
INFEKSI HEPATITIS C**

LAPORAN SCIENTIST

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

NADIA MARTHA LEFI

22.11.5140

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

JALUR SCIENTIST

**ANALISIS OPTIMASI HYPERPARAMETER DAN FEATURE SELECTION
PADA MODEL XGBOOST UNTUK PREDIKSI INFEKSI HEPATITIS C**

yang disusun dan diajukan oleh

Nadia Martha Lefi

22.11.5140

telah disetujui oleh Dosen Pembimbing

pada tanggal 17 Desember 2025

Dosen Pembimbing,



Majid Bahardj, S.Kom., M.Eng
NIK. 190102393

HALAMAN PENGESAHAN

JALUR SCIENTIST

**ANALISIS OPTIMASI HYPERPARAMETER DAN FEATURE SELECTION
PADA MODEL XGBOOST UNTUK PREDIKSI INFEKSI HEPATITIS C**

yang disusun dan diajukan oleh

Nadia Martha Lefi
22.11.5140

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ferlan Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276

Uyock Anggoro Saputro, S.Kom., M.Kom.
NIK. 190302419

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nadia Martha Lefi

NIM : 22.11.5140

Menyatakan bahwa Laporan dengan judul berikut:

Analisis Optimasi Hyperparameter dan Feature Selection pada Model XGBoost untuk Prediksi Infeksi Hepatitis C

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 17 Desember 2025

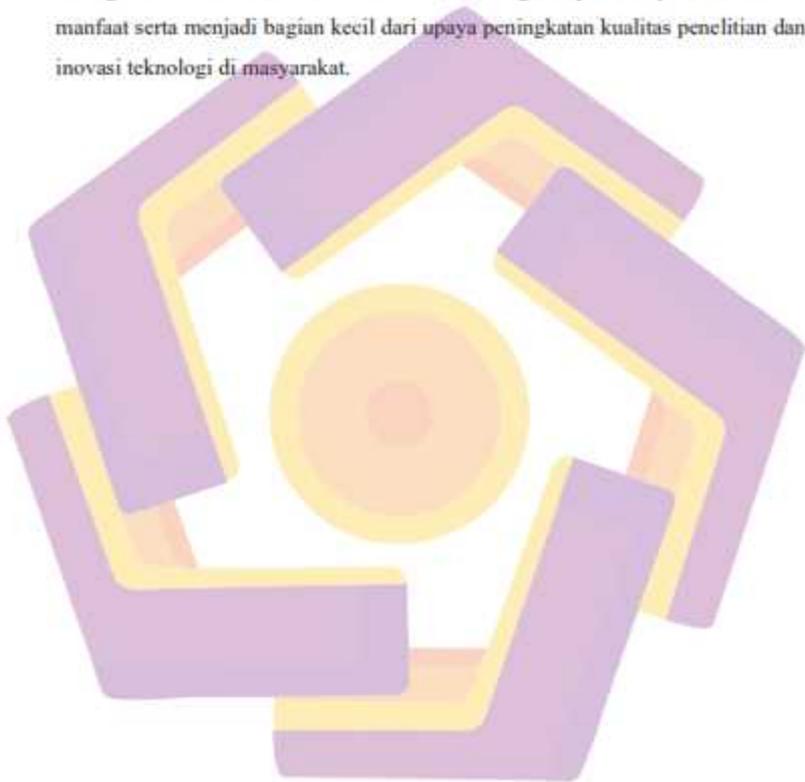
Yang Menyatakan,



Nadia Martha Lefi

HALAMAN PERSEMBAHAN

Laporan ini penulis persembahkan kepada Universitas Amikom Yogyakarta sebagai bentuk kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang analisis data dan kecerdasan buatan. Semoga karya ini dapat memberikan manfaat serta menjadi bagian kecil dari upaya peningkatan kualitas penelitian dan inovasi teknologi di masyarakat.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga laporan non-reguler ini dapat diselesaikan dengan baik. Penyusunan laporan ini tidak terlepas dari dukungan, bimbingan, dan bantuan dari berbagai pihak yang telah memberikan kontribusi besar selama proses penelitian hingga terselesaikannya laporan ini. Oleh karena itu, dengan penuh hormat, penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Majid Rahardi, S.Kom., M.Eng., selaku Dosen Pembimbing, atasa bimbingan, arahan, serta dukungan yang diberikan kepada penulis selama proses pengerjaan penelitian ini.
2. Ibu Yumariyah, S.Pd.I, selaku ibu kandung penulis, atas kasih sayang, doa, dan dukungan yang tiada henti yang selalu menjadi sumber kekuatan dalam menyelesaikan laporan ini.
3. Fransisko Andrade Laiskodat, atas dukungan, kerja sama, serta senantiasa membantu dan menemani penulis dalam menyelesaikan laporan ini.
4. Almaas dan Cokorda, selaku teman penulis, atas dukungan dan bantuan selama proses menyelesaikan laporan ini.

Penulis menyadari bahwa laporan ini masih memiliki keterbatasan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat bagi pembaca serta berkontribusi pada pengembangan ilmu pengetahuan di bidang machine learning dan kesehatan.

Yogyakarta, 20 November 2025

Penulis

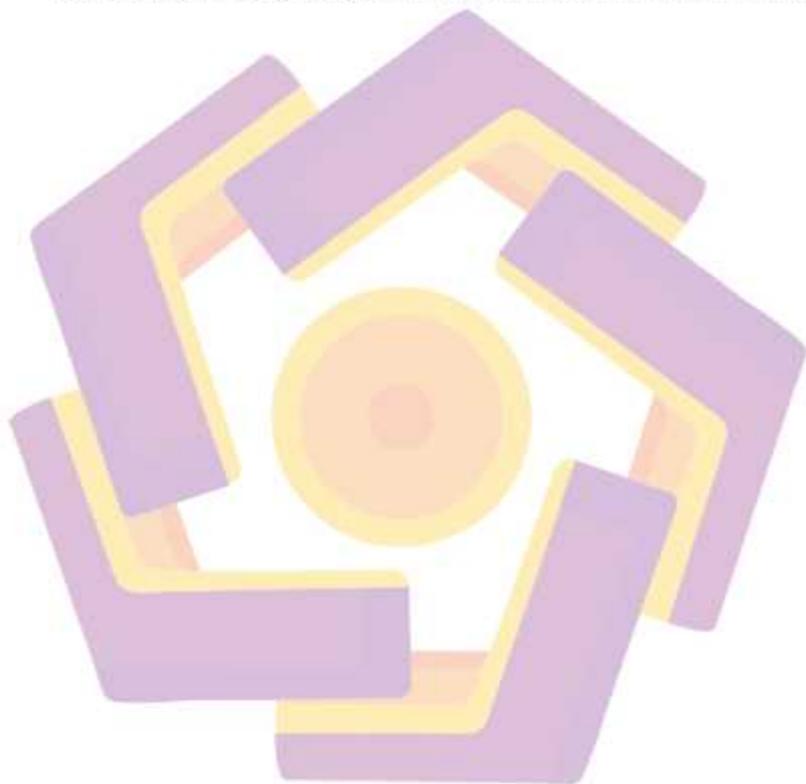
DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan.....	iii
Halaman Pernyataan Keaslian Karya.....	iv
Halaman Persembahan.....	v
Kata Pengantar.....	vi
Daftar Isi.....	vii
Daftar Tabel.....	ix
Daftar Gambar.....	x
Daftar Lampiran.....	xi
Daftar Lambang dan Singkatan.....	xii
Daftar Istilah.....	xiii
Intisari.....	xv
<i>Abstract</i>	xvi
Bab I Pendahuluan.....	1
1.1 Gambaran Umum.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan.....	3
Bab II Tinjauan Pustaka.....	4
2.1 Studi Literatur.....	4
2.2 Landasan Teori.....	6
2.2.1 Hepatitis C.....	6
2.2.2 Machine Learning.....	7
2.2.3 Extreme Gradient Boosting (XGBoost).....	7
2.2.4 Feature Selection.....	8
2.2.5 Hyperparameter Tuning.....	8
2.2.6 Integrasi Feature Selection dan Hyperparameter Tuning.....	8
2.2.7 Synthetic Minority Over-Sampling Technique (SMOTE).....	9
2.2.8 Metrik Evaluasi.....	9

BAB III Metode Penelitian	10
3.1 Metode	10
3.1.1. Studi Literatur	11
3.1.2. Akuisisi Data	11
3.1.3. Preprocessing	11
3.1.4. Split Data	12
3.1.5. Balancing Data	12
3.1.6. Seleksi Fitur	13
3.1.7. Modelling	13
3.1.8. Optimisasi Hyperparameter	14
3.1.9. Evaluasi Model	15
BAB IV Pembahasan	18
4.1. Sub Pembahasan	18
4.1.1. Akuisisi Data	18
4.1.2. Split Data	20
4.1.3. Balancing Data	20
4.1.4. Hasil Evaluasi Model	21
4.1.5. Visualisasi Hasil Evaluasi	23
4.1.6. Analisis Per Model	26
BAB V Kesimpulan	29
4.3. Kesimpulan	29
4.4. Saran	30
Referensi	31
Curriculum Vitae	34
Lampiran dan Bukti Pendukung	35
a. Letter of Acceptance (LOA)	35
b. Lembar Review	36
c. Bukti Terbit/Terindex	38
d. Bukti pembayaran	38

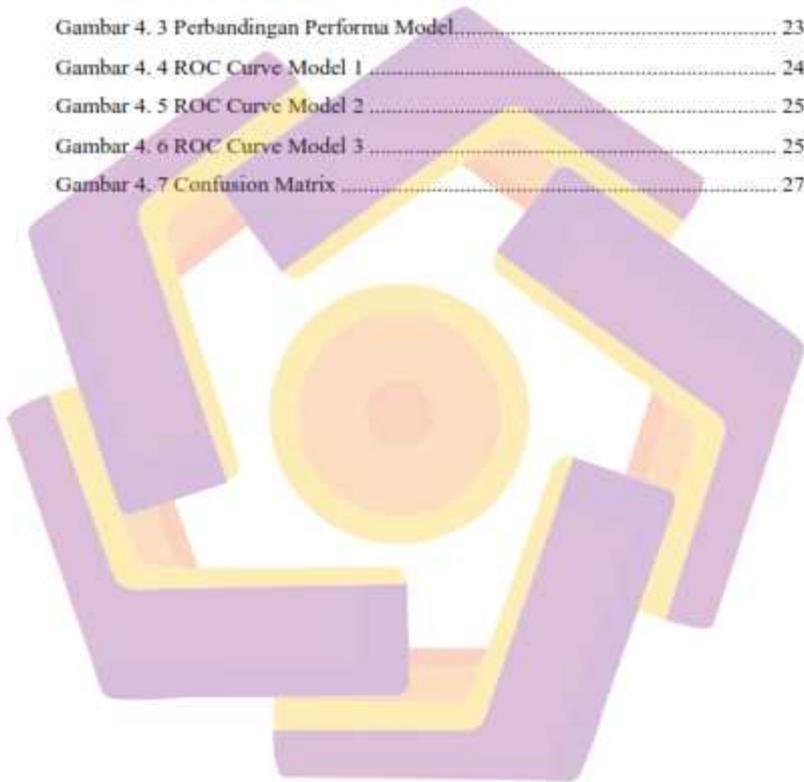
DAFTAR TABEL

Tabel 3. 1 Skema Evaluasi Model.....	17
Tabel 4. 1 Contoh Data Mentah.....	19
Tabel 4. 2 Metrik Evaluasi Model.....	22



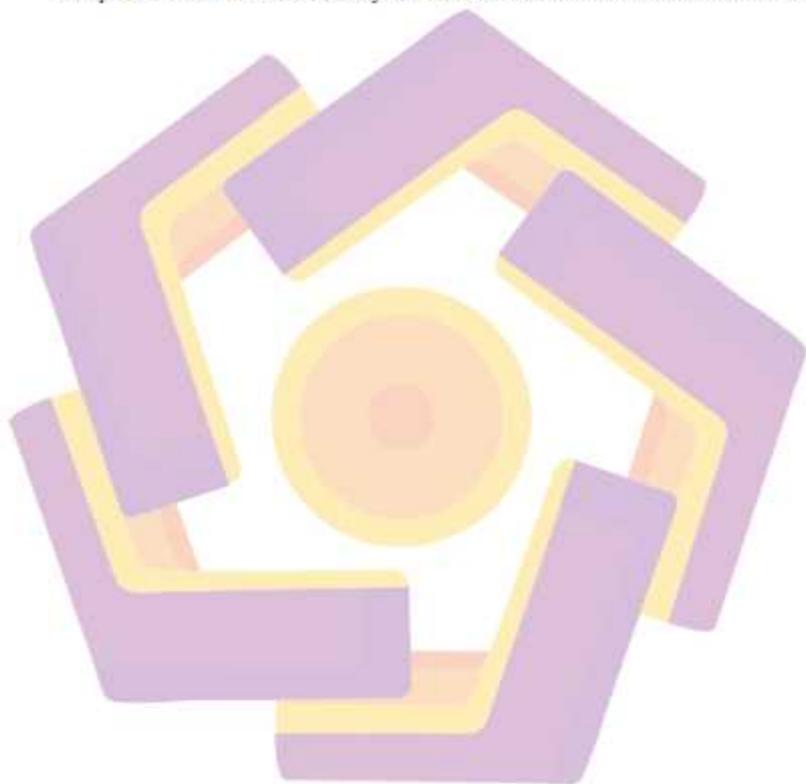
DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	10
Gambar 4. 1 Distribusi kelas sebelum dilakukan SMOTE.....	18
Gambar 4. 2 Distribusi kelas setelah dilakukan SMOTE	21
Gambar 4. 3 Perbandingan Performa Model.....	23
Gambar 4. 4 ROC Curve Model 1	24
Gambar 4. 5 ROC Curve Model 2	25
Gambar 4. 6 ROC Curve Model 3	25
Gambar 4. 7 Confusion Matrix	27



DAFTAR LAMPIRAN

Lampiran 1 Bukti Letter of Acceptance (LOA)	35
Lampiran 2 Bukti Terbit pada Journal of Applied Informatics and Computing ...	38
Lampiran 3 Bukti Instruksi Pembayaran dan Nominal	38



DAFTAR LAMBANG DAN SINGKATAN



HCV	Virus Hepatitis C
ML	<i>Machine Learning</i>
XGBoost	Extreme Gradient Boosting
ROC	<i>Receiver Operating Characteristic</i>
AUC	<i>Area Under Curve</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
SMOTE	Synthetic Minority Over-Sampling Technique
ANOVA	Analysis of Variance

DAFTAR ISTILAH

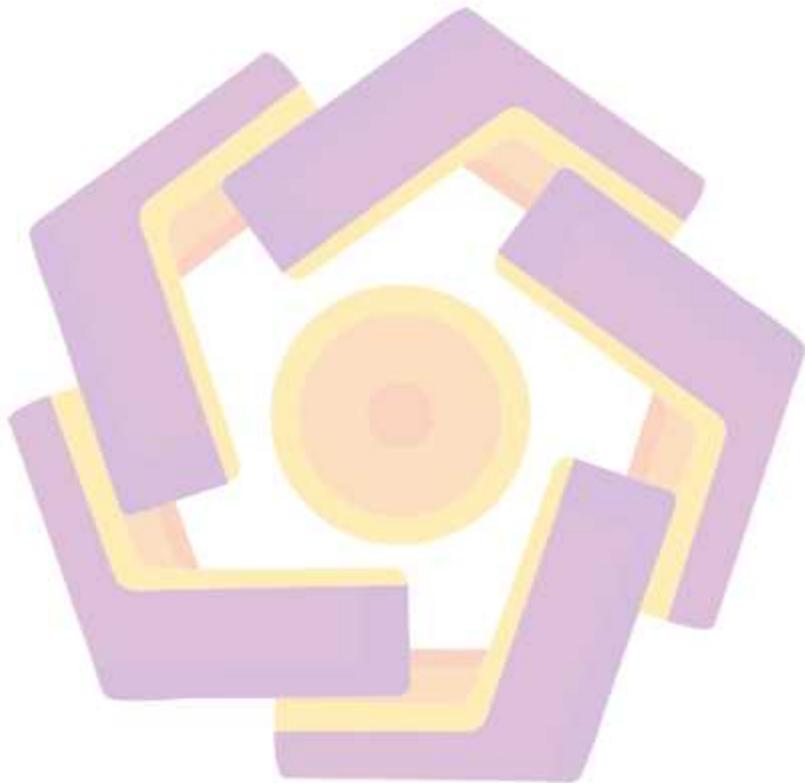
Hepatitis C	Penyakit menular yang menyerang hati dan dapat berkembang menjadi kondisi kronis.
Dataset	Kumpulan data analisis.
<i>Preprocessing</i>	Tahapan pembersihan dan persiapan data sebelum digunakan model.
<i>Missing Value</i>	Nilai kosong yang harus ditangani agar data konsisten.
<i>Encoding</i>	Mengubah data kategorikal menjadi bentuk numerik.
Normalisasi	Penyeragaman skala nilai fitur.
<i>Split Data</i>	Membagi dataset menjadi data latih dan data uji.
<i>Balancing Data</i>	Menyeimbangkan jumlah sampel antar kelas untuk mengatasi ketidakseimbangan kelas.
<i>Oversampling</i>	Teknik menambah sampel pada kelas minoritas.
<i>Feature Selection</i>	Proses memilih fitur paling relevan untuk meningkatkan performa model.
<i>ANOVA F-score</i>	Metode statistik untuk menilai relevansi fitur terhadap label.
<i>Hyperparameter Tuning</i>	Penyesuaian parameter model untuk memperoleh performa terbaik.
GridSearchCV	Metode pencarian kombinasi hyperparameter optimal menggunakan validasi silang.
<i>Training Set</i>	Data yang digunakan untuk melatih model.
<i>Confusion Matrix</i>	Tabel evaluasi yang menunjukkan prediksi benar dan salah.
<i>Accuracy</i>	Proporsi prediksi benar terhadap total sampel.
<i>Precision</i>	Tingkat ketepatan prediksi positif.
<i>Recall</i>	Kemampuan model menangkap seluruh kasus positif.
<i>F1-score</i>	Rata-rata harmonik precision dan recall.

ROC Curve

Kurva yang menunjukkan hubungan antara TPR dan FPR.

Macro Average

Rata-rata metrik yang memberi bobot sama untuk setiap kelas.



INTISARI

Hepatitis C merupakan penyakit hati yang dapat berkembang menjadi kondisi kronis seperti sirosis dan kanker hati. Deteksi dini sangat penting dilakukan dan dapat didukung melalui pendekatan machine learning. Penelitian ini menganalisis pengaruh seleksi fitur dan optimasi hyperparameter terhadap performa model XGBoost dalam mengklasifikasikan infeksi Hepatitis C. Dataset yang digunakan berasal dari Kaggle dan berisi atribut hasil uji laboratorium. Tahap preprocessing meliputi penanganan nilai hilang, pengkodean variabel kategorikal, penghapusan kelas pencilan, serta normalisasi data menggunakan StandardScaler. Setelah dilakukan pembagian data secara stratifikasi, data latih diseimbangkan menggunakan teknik SMOTE. Seleksi fitur dilakukan dengan metode ANOVA F-score, sedangkan hyperparameter tuning dilakukan menggunakan GridSearchCV. Penelitian ini membandingkan tiga skenario model, yaitu baseline, model dengan seleksi fitur, serta model dengan kombinasi seleksi fitur dan hyperparameter tuning. Hasil evaluasi menunjukkan bahwa model ketiga memberikan performa terbaik dengan akurasi 96%, presisi 79%, recall 81%, dan F1-score 78%, meskipun terjadi sedikit penurunan pada nilai ROC AUC. Pendekatan ini terbukti efektif dalam meningkatkan performa model dan relevan untuk mendukung sistem diagnosis Hepatitis C yang lebih akurat.

Kata kunci: Klasifikasi, Seleksi Fitur, Hepatitis C, Optimasi Hyperparameter, XGBoost.

ABSTRACT

Hepatitis C is a liver disease that can progress to chronic conditions such as cirrhosis and liver cancer. Early detection is essential and can be supported through machine learning approaches. This study analyzes the effect of feature selection and hyperparameter tuning on the performance of the XGBoost model in classifying hepatitis C infection. The dataset, obtained from Kaggle, contains laboratory test attributes. The preprocessing stage involved handling missing values, encoding categorical variables, removing outlier classes, and normalizing data using StandardScaler. After stratified splitting, the training set was balanced using the SMOTE technique. Feature selection was carried out using the ANOVA F-score method, and hyperparameter tuning was performed using GridSearchCV. Three model scenarios were compared: baseline, with feature selection, and with combined feature selection and hyperparameter tuning. The evaluation results showed that the third model achieved the best performance with 96% accuracy, 79% precision, 81% recall, and a 78% F1-score, despite a slight decrease in the ROC AUC value. This approach has proven effective in improving model performance and is relevant for supporting more accurate hepatitis C diagnosis systems.

Keyword: *Classification, Feature Selection, Hepatitis C, Hyperparameter Optimization, XGBoost.*