

KLASIFIKASI DETEKSI PASIEN HEPATITIS C VIRUS (HCV) MENGUNAKAN METODE ARTIFICIAL NEURAL NETWORK(ANN) DAN OVERSAMPLING SMOTE

Maylinna Rahayu Ningsih¹, Ahmad Ubai Dullah², Apri Dwi Lestari³, Jumanto⁴, Alamsyah⁵, Much Aziz Muslim⁶

Departemen Ilmu Komputer/Fakultas Matematika dan Ilmu Pengetahuan Alam/Program Studi Teknik Informatika
Universitas Negeri Semarang
Semarang, Indonesia

e-mail: maylinnarahayuningsih@students.unnes.ac.id¹, ubaid@students.unnes.ac.id², apriidwi@students.unnes.ac.id³
jumanto@mail.unnes.ac.id⁴, alamsyah@mail.unnes.ac.id⁵, a212muslim@mail.unnes.ac.id⁶

Abstrak

Virus Hepatitis C (HCV) merupakan masalah kesehatan global dengan lebih dari 160 juta orang terinfeksi setiap tahunnya dan menyebabkan lebih dari 350 ribu kematian. HCV merupakan penyakit yang mengacu pada peradangan hati *Crossover Hepatitis C* berkembang pada 50-80% pasien, bahkan secara global pada awal tahun 2020 sekitar 56,8 juta orang terinfeksi virus ini. Banyak penelitian yang telah dilakukan, namun penelitian-penelitian sebelumnya belum mendapatkan hasil yang optimal dalam klasifikasi pasien hepatitis C. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi deteksi pasien Hepatitis C dengan menggunakan metode *Artificial Neural Network* (ANN). Model ini terdiri dari beberapa lapisan neuron buatan yang terhubung, masing-masing dengan nilai bobot dan bias. Lapisan pertama dan terakhir masing-masing adalah variabel input dan *output*, dan lapisan perantara terdiri dari inferensi matematis dan transformasi parameter input menjadi prediksi di neuron tersembunyi. Pada penelitian ini, data diolah melalui beberapa tahapan, seperti menghilangkan missing value, mengubah nilai kategori menjadi numerik, dan melakukan oversampling dengan metode SMOTE. Metode ini digunakan karena label Category pada dataset tidak seimbang sehingga digunakan metode untuk menyeimbangkan distribusi kelas pada kumpulan dataset. Model ANN yang diusulkan terdiri dari dua lapisan, input dan *output*, dengan aktivasi ReLU dan softmax. Hasil evaluasi menunjukkan bahwa model ANN mencapai akurasi 97% untuk data pelatihan dan pengujian. Hasil performa model menunjukan penggunaan model ANN dan SMOTE berhasil meningkatkan kinerja dan mengurangi kesalahan klasifikasi.

Kata kunci: Hepatitis C Virus(HCV), Deteksi, Artificial Neural Network(ANN), SMOTE, Neural Network

Abstract

Hepatitis C virus (HCV) is a global health problem with more than 160 million people infected annually and causes more than 350 thousand deaths. HCV is a disease that refers to the inflammation of the liver. Cross-over Hepatitis C develops in 50-80% of patients, even globally in early 2020 about 56.8 million people are infected with this virus. Many studies have been conducted, but previous studies have not obtained optimal results in the classification of hepatitis C patients. This study aims to improve the classification accuracy of Hepatitis C patient detection by using the Artificial Neural Network (ANN) method. This model consists of several layers of connected artificial neurons, each with a weight and bias value. The first and last layers are input and output variables respectively, and the intermediate layer consists of mathematical inference and transformation of input parameters into predictions in the hidden neurons. In this study, the data is processed through several stages, such as removing missing values, converting category values to numeric, and oversampling with the SMOTE method. This method is used because the Category labels in the dataset are not balanced so a method is used to balance the class distribution in the dataset set. The proposed ANN model consists of two layers, input and output, with ReLU and softmax activations. The evaluation results showed that the ANN model achieved 97% accuracy for both training and testing data. The model performance results indicate that the use of ANN and SMOTE models can improve performance and reduce classification errors.

Keywords: Hepatitis C Virus(HCV), Detection, Artificial Neural Network(ANN), SMOTE, Neural network

I. PENDAHULUAN

Setiap tahun lebih dari 160 juta orang terinfeksi oleh virus Hepatitis C atau HCV, sebanyak 350.000 orang meninggal dunia akibat HCV. HCV merupakan penyakit yang mengacu pada peradangan hati *Crossover Hepatitis C* berkembang [1] pada 50-80% pasien, bahkan secara global pada awal tahun 2020 sekitar 56,8 juta orang terinfeksi virus ini [2]. Konsekuensi dari penyakit ini memicu beberapa penyakit lainnya diantaranya sirosis hati, kanker hati, hepatoseluler, gagal hati, dan kematian [3], [4]. Penyakit HCV merupakan penyakit yang ditularkan melalui darah seperti peralatan kesehatan yang tidak aman, praktek seksual, tranfusi darah, penggunaan narkoba, atau paparan perinatal [5]. Penyakit ini memicu kerusakan hati yang bervariasi dalam kurun waktu 6 bulan, dimana tingkat peradangan pada hari bisa akut maupun kronik [5], [6]. Penyakit hepatitis C ini merupakan penyakit berbahaya [7] yang berkembang perlahan hingga menyebabkan penyakit kanker.

Pada era moderen dimana teknologi berkembang dalam segala aktivitas manusia, komputer telah banyak dilatih menggunakan bahasa pemrograman yang dapat belajar dari data untuk melakukan tugas seperti membuat prediksi, mengenali pola dan sebagainya. *Machine learning* merupakan ilmu pengembangan algoritma dan model yang mampu digunakan dalam beberapa kasus. Penelitian terkait penyakit HCV menggunakan *machine learning* tentunya sudah ada sebelumnya. Seperti pada penelitian [8] yang melakukan prediksi penyakit Hepatitis C menggunakan algoritma *Logistic Regression* dengan akurasi 90,24%. Penelitian lainnya oleh [5] menggunakan *Cascade RF-LR* (dengan SMOTE) yang membuktikan penggunaan SMOTE bisa meningkatkan performa model menjadi 96,38%. Penelitian lainnya yang menggunakan *protein interaction networks* (PINs) [9] untuk hepatitis B (HBV) dan hepatitis C (HCV) dan hasilnya penerapan prosedur peneringkatan statistik jaringan di antaranya lebih penting. Penelitian [10] menggunakan *Transfer Learning and Hybrid Quantum Neural Networks* (QNNs) berhasil menerapkan metode *neural network* jenis *Hybrid Quantum* dengan mendapatkan akurasi 93,75%

Perkembangan teknologi, *computed Tomography* (CT) yang merupakan teknik pencitraan yang banyak digunakan untuk mendeteksi lesi hati dan kelainan lainnya terkait dengan infeksi HCV bisa digunakan. Namun, interpretasi CT dapat menjadi tantangan, memakan waktu, dan tunduk pada variabilitas antar pengamat, sehingga menyulitkan ahli radiologi untuk mendiagnosis HCV secara akurat. Perkembangan itu juga diikuti dengan penelitian dibidang *machine learning* seperti yang disebutkan sebelumnya. Namun, penelitian-penelitian sebelumnya belum mendapatkan hasil yang optimal dalam klasifikasi pasien hepatitis C.

Dari paparan sebelumnya, penelitian ini mengusulkan penggunaan algoritma *Artificial neural network* (ANN) dan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengklasifikasikan penyakit virus Hepatitis C atau HCV. Kontribusi dalam penelitian ini ditujukan untuk meningkatkan performa model sehingga menghasilkan akurasi yang lebih baik dengan metode yang diusulkan.

II. TINJAUAN PUSTAKA

1. Virus Hepatitis C atau HCV

Virus hepatitis C (HCV) adalah virus RNA beruntai tunggal yang diklasifikasikan ke dalam delapan genotipe utama, yaitu genotipe 1 hingga 8 [11]. HCV dianggap sebagai ancaman global yang utama dan dapat menyebabkan berbagai komplikasi termasuk sirosis hati, karsinoma hepatoseluler, dan kematian. Pilihan pengobatan untuk HCV sebelumnya didasarkan pada interferon alfa pegilasi, yang dikaitkan dengan efek samping neuropsikiatri, dan hal ini berkontribusi pada pengecualian pasien dengan SMI dari pengobatan HCV, program eliminasi, dan uji klinis. Selain itu, asumsi kepatuhan yang buruk, akses ke layanan kesehatan yang minim, serta stigma dan kerentanan populasi ini muncul sebagai hambatan dan berkontribusi pada rendahnya tingkat pengobatan dan kemanjuran [12].

2. SMOTE

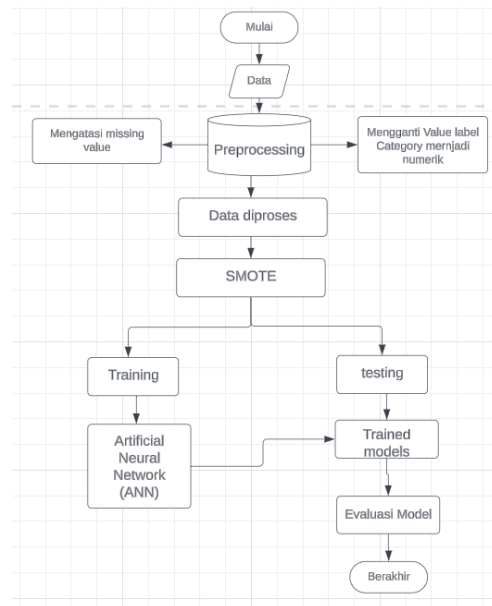
Synthetic Minority Over-sampling Technique (SMOTE) adalah metode yang banyak digunakan untuk mengatasi ketidakseimbangan kelas dengan menghasilkan sampel sintetis untuk kelas minoritas [13]. SMOTE menghasilkan sampel sintetis dengan memilih sampel acak dari kelas minoritas, mengidentifikasi k-tetangga terdekat, dan melakukan interpolasi antara sampel yang dipilih dan salah satu tetangganya [14].

3. *Artificial neural network* (ANN)

Artificial Neural Network (ANN), bagian dari model ML, telah terbukti berhasil dalam analisis deret waktu, menangkap hubungan antara variabel input dan *output* untuk memodelkan sistem yang kompleks [15]. ANN dapat memproses beragam input, menghasilkan banyak prediksi, dan menangani kumpulan data yang besar, sehingga menunjukkan keefektifan dalam meningkatkan kinerja mesin dan mengurangi emisi.

III. ANALISIS PENELITIAN

Alur desain algoritma ANN yang diusulkan digambarkan seperti yang ditunjukkan pada Gambar 1. Setiap proses pada Gambar 1 akan dijelaskan secara rinci pada bagian selanjutnya.



Gambar 1. Flowchart Diagram Untuk Metode Yang Diusulkan

1. Data Processing (Pemrosesan Data)

Data preprocessing secara sederhana merupakan prosedur dalam menyiapkan data [16] hingga siap digunakan untuk proses selanjutnya. Ada beberapa langkah yang terlibat dalam memproses data, sebagai berikut:

a. Mengatasi Missing Value

Missing value pada model algoritma ANN merupakan data yang hilang atau tidak lengkap sehingga dapat mempengaruhi kinerja model karena model tidak dapat memproses data yang tidak lengkap seperti pada Gambar 2. Oleh karena itu, perlu dilakukan penghilangan missing value dengan cara menghapus kolom yang berisi missing value.

Unnamed: 0	Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT	
0	1	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	2	0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	3	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	4	0=Blood Donor	32	m	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	5	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7

(a)

Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT	
0	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	0=Blood Donor	32	m	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7

(b)

Gambar 2. (a) Dataset Sebelum Diatasi, (b) Dataset Setelah Diatasi Missing Value

Tujuan dari menghilangkan Missing Value adalah untuk memudahkan model ANN dalam melakukan proses klasifikasi dan data yang tidak diperlukan akan dibuang atau dihapus, sehingga lebih efisien dan optimal. Tabel dengan nama “Unnamed” dihapus dan beberapa missing value pada fitur dihapus agar lebih optimal dalam memproses dataset yang diberikan.

b. Mengganti nilai Kategori menjadi angka

Mengganti nilai Kategori menjadi angka adalah proses dalam algoritma ANN yang mengubah nilai kategorikal dalam data menjadi nilai angka untuk memudahkan pemrosesan oleh model ANN. Hal ini dikarenakan model ANN membutuhkan data numerik untuk diproses dan dipelajari. Namun, pada data medis, terdapat banyak variabel kategorikal seperti jenis kelamin, jenis obat, dan lain sebagainya. Oleh karena itu, sebelum data medis dapat diproses oleh model ANN, nilai kategorikal dari variabel tersebut perlu diubah menjadi nilai numerik. Sebagai contoh, pada Gambar 3, variabel kategorikal “donor darah”, “Hepatitis”, “Fibrosis”, dan “Chirrosis” diubah menjadi nilai numerik “0”, “1”, “2”, dan “3”.

Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT	
0	0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	0=Blood Donor	32	m	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	0=Blood Donor	32	m	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7

(a)

Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT	
0	0	32	1	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106.0	12.1	69.0
1	0	32	1	38.5	70.3	18.0	24.7	3.9	11.17	4.80	74.0	15.6	76.5
2	0	32	1	46.9	74.7	36.2	52.6	6.1	8.84	5.20	86.0	33.2	79.3
3	0	32	1	43.2	52.0	30.6	22.6	18.9	7.33	4.74	80.0	33.8	75.7
4	0	32	1	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76.0	29.9	68.7

(b)

Gambar 3. (a) Dataset Sebelum Mengganti Category, (b) Dataset Setelah Mengganti Category

Model algoritma yang digunakan hanya dapat memproses data numerik. Oleh karena itu, data yang bukan berupa angka akan diubah menjadi angka. Sebagai contoh, pada Gambar 3, variabel kategori “donor darah”, “Hepatitis”, “Fibrosis”, dan “Chirrosis” diubah menjadi nilai numerik “0”, “1”, “2”, dan “3”. Dan juga mengubah variabel jenis kelamin “pria” dan “wanita” menjadi bentuk numerik untuk memudahkan pemrosesan. Variabel-variabel tersebut diubah menjadi “1” dan “0”. Serta variabel jenis kelamin “pria” dan “wanita” diubah menjadi bentuk numerik agar lebih mudah diproses. Diubah menjadi “1” dan “0”.

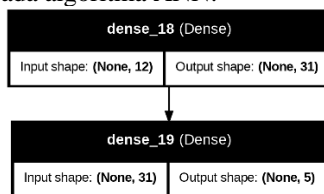
2. *Oversampling* data menggunakan metode *Synthetic Minority Oversampling Technique(SMOTE)*

Metode ini digunakan karena label *Category* pada dataset tidak seimbang sehingga digunakan metode untuk menyeimbangkan distribusi kelas pada kumpulan dataset. Dalam masalah klasifikasi, distribusi kelas dikatakan tidak seimbang ketika *jumlah* sampel di satu kelas secara signifikan lebih rendah daripada jumlah sampel di kelas lainnya. Metode SMOTE menghasilkan sampel sintesis dari kelas minoritas dengan membuat sampel baru yang menginterpolasi antara sampel yang ada dari kelas minoritas [17]–[19]. Hal ini dilakukan dengan memilih sampel acak dari kelas minoritas dan menemukan k tetangga terdekatnya dalam ruang fitur. Sampel sintesis kemudian dibuat dengan memilih secara acak salah satu dari k-tetangga terdekat dan melakukan interpolasi di antara dua sampel.

Setelah dilakukan *oversampling* data dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Dataset yang dibagi adalah x dan y, dengan ukuran data *testing* sebesar 0.3 atau 30% dari keseluruhan dataset. Hasil dari pemisahan data latih dan data uji adalah empat variabel yaitu x_train, x_test, y_train, dan y_test, yang masing-masing *merepresentasikan* subset data latih dan data uji untuk variabel independen dan dependen. Data berhasil dibagi menjadi empat variabel, dan ukuran (*shape*) dari setiap subset data dapat dilihat dengan menggunakan metode “.shape”. Variabel x_train.shape, x_test.shape, y_train.shape, dan y_test.shape akan menghasilkan sebuah tuple yang merepresentasikan dimensi atau ukuran dari masing-masing subset data.

3. Pemodelan

Model Program menggunakan library Keras untuk membuat model *Neural Network* (NN) dengan 2 layer, yaitu 1 layer *input* dan 1 layer *output* [20], [21] pustaka Keras dibuat untuk melatih, dan menguji model *Neural Network* (NN) [22], [23]. *Artificial neural network* (ANN) merupakan algoritma berupa model pembelajaran mesin yang terinspirasi dari cara kerja otak *manusia* [16]–[18]. Model ini terdiri dari beberapa lapisan neuron buatan yang terhubung, masing-masing dengan nilai bobot dan bias. Lapisan pertama dan terakhir masing-masing adalah variabel input dan *output*, dan lapisan perantara terdiri dari inferensi matematis dan transformasi parameter input menjadi prediksi di neuron tersembunyi [19]. ANN mempelajari batasan dan prediksi dengan cara yang mirip dengan sistem biologis otak manusia, contoh yang sering digunakan adalah MLP atau *Multi-Layer Perceptron*. Gambar 4. Menunjukkan model *layer* yang digunakan pada algoritma ANN.



Gambar 4. *Layer* yang digunakan pada ANN

4. Evaluasi Model

Proses menggunakan model yang telah dilatih untuk memprediksi *output* untuk data input baru. Dalam ANN, model terdiri dari *serangkaian* lapisan yang menerima input dan menghasilkan *output*. Setelah model dilatih dengan data pelatihan, model tersebut dapat digunakan untuk memprediksi *output* untuk data yang belum pernah dilihat sebelumnya.

Hasilnya akan ditampilkan dalam bentuk hasil prediksi dari akurasi yang telah ditentukan dari data latih dan data uji. Semua *metrik* tersebut dihitung dengan rumus di bawah ini:

$$accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{1}$$

$$precision = TP/(TP + FP) \tag{2}$$

$$recall = TP/(TP + FN) \tag{3}$$

$$F1\ Score = \frac{2 \times (recall \times precision)}{(recall + precision)} \tag{4}$$

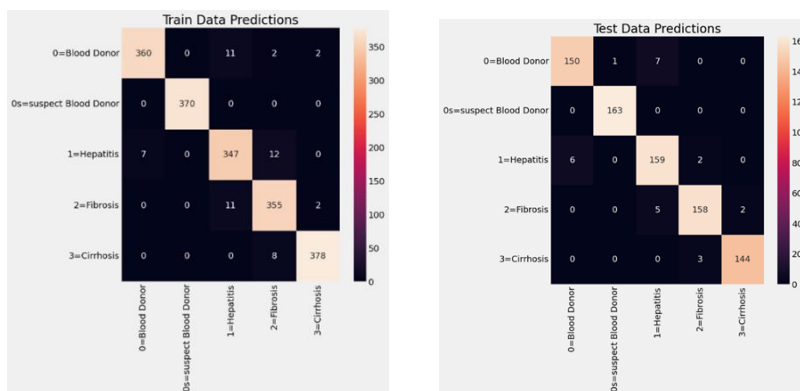
Dengan TP adalah *True Positive*, jumlah data positif yang diprediksi positif, TN adalah *True Negative*, jumlah data negatif yang diprediksi negatif, FP adalah *False Positive*, jumlah data positif yang diprediksi negatif, dan FN adalah *False Negative*, jumlah data negatif yang diprediksi positif.

IV. HASIL DAN PEMBAHASAN

Bagian ini menyajikan temuan-temuan utama dari penelitian ini, bersama dengan analisis data dan interpretasi hasil untuk menjawab pertanyaan-pertanyaan penelitian. Hasil akurasi data *training* dan test yang diperoleh dari penggunaan algoritma ANN dengan menggunakan *hard* model dengan jumlah lapisan keras sebanyak 31 dan 9, memberikan hasil seperti yang ditunjukkan pada Gambar 5 dan Tabel 1.

TABEL I
 HASIL PREDIKSI PADA DATA TRAIN

	precision	recall	f1-score	Support
0	0.96	0.95	0.96	158
1	0.99	1.00	1.00	163
2	0.93	0.95	0.94	167
3	0.7	0.96	0.96	165
4	0.99	0.98	0.98	147
Accuracy			0.97	800
Macro avg	0.97	0.97	0.97	800
Weighted avg	0.97	0.97	0.97	800



Gambar 5 . Hasil *Confusion Matrix* prediksi pada *Data Train* dan *Test*

TABEL II
 HASIL PREDIKSI PADA DATA TRAIN

	precision	recall	f1-score	Support
0	0.96	0.95	0.96	158
1	0.99	1.00	1.00	163
2	0.93	0.95	0.94	167
3	0.7	0.96	0.96	165
4	0.99	0.98	0.98	147
Accuracy			0.97	800
Macro avg	0.97	0.97	0.97	800
Weighted avg	0.97	0.97	0.97	800

Kasus ini, model ANN berkinerja cukup baik, dengan akurasi 0,97. Rata-rata makro dan rata-rata tertimbang juga menunjukkan bahwa model memiliki kinerja yang sama di setiap kelas, dan tidak ada bias yang signifikan terhadap kelas tertentu. Data latih yang digunakan adalah 1865 dari total dataset atau sekitar 69.9%. Hasil akurasi yang memiliki nilai 97% untuk akurasi menunjukkan bahwa akurasi ini lebih baik atau lebih optimal dibandingkan dengan algoritma pada penelitian sebelumnya. Berdasarkan hasil *Classification Report* pada Tabel, model ANN memiliki kinerja yang baik dengan akurasi 0.97 atau 97%. *Precision*, *recall*, dan *f1-score* masing-masing merepresentasikan seberapa baik model ANN dapat memprediksi kelas yang benar, baik secara keseluruhan maupun untuk setiap kelas secara individual.

1. Pembahasan

Berdasarkan hasil yang telah dijelaskan sebelumnya, model ANN yang dibangun mampu mengklasifikasikan pasien hepatitis C dengan akurasi sebesar 97% pada data *training* dan 97% pada data *testing*. Hal ini menunjukkan bahwa model ANN memiliki performa yang baik dalam mengklasifikasikan pasien hepatitis C.

TABEL III
PERBANDINGAN HASIL PENELITIAN

Author	Algorithm	Accuracy
Md. Julker Nayeem, et al. [24]	Random Forest	92.41%
Li, H. J. Chiu, et al. [5]	Cascade RF-LR (dengan SMOTE)	96,38%
Simos, et al. [9]	<i>protein interaction networks</i> (PINs)	93,75%
Metode yang diusulkan	Artificial Neural Network	97%

Berdasarkan hasil perbandingan pada tabel 3. dapat dilihat bahwa hasil akurasi yang didapatkan dengan menggunakan metode yang diusulkan memiliki hasil yang lebih tinggi dibandingkan dengan metode-metode sebelumnya.

V. KESIMPULAN

Model *Artificial Neural Network* (ANN) dapat digunakan untuk mengklasifikasikan dataset hepatitis C dengan akurasi yang baik. Pada percobaan pertama dengan menggunakan 70% untuk data *training* dan 30% untuk data *testing*, didapatkan hasil akurasi sebesar 97% untuk data *training* dan 97% untuk data *testing*. Hasil ini menunjukkan bahwa model mampu memprediksi dengan baik pada data yang belum pernah dilihat sebelumnya. Sedangkan hasil *confusion matrix* plot menunjukkan bahwa model ANN mampu mengenali dengan benar sebagian besar data pada dataset, namun masih terdapat beberapa data yang salah klasifikasi. Dari hasil tersebut, dapat disimpulkan bahwa model ANN dapat digunakan sebagai alternatif untuk mengklasifikasikan dataset hepatitis C. Namun, diperlukan pengembangan lebih lanjut untuk meningkatkan performa model dan mengurangi jumlah data yang salah klasifikasi. Selain itu, penelitian lebih lanjut juga dapat dilakukan untuk membandingkan performa model ANN dengan model klasifikasi lainnya.

REFERENSI

- [1] Institute of Electrical and Electronics Engineers. Bangladesh Section, IEEE Region 10, dan Institute of Electrical and Electronics Engineers, 2020 IEEE Region 10 Symposium (TENSYP): 5-7 June 2020, Dhaka, Bangladesh. .
- [2] M.-H. Hsieh, T.-Y. Kao, dan T.-H. Hsieh, "Predictors of liver fibrosis changes assessed by paired liver biopsies in chronic hepatitis C patients treated with direct-acting antivirals," *J. Microbiol. Immunol. Infect.*, 2024, doi: <https://doi.org/10.1016/j.jmii.2024.08.005>.
- [3] A. Petruzzello, S. Marigliano, G. Loquercio, A. Cozzolino, dan C. Cacciapuoti, "Global epidemiology of hepatitis C virus infection: An up-date of the distribution and circulation of hepatitis C virus genotypes," *World Journal of Gastroenterology*, vol. 22, no. 34. Baishideng Publishing Group Co, hal. 7824–7840, Sep 2016, doi: 10.3748/wjg.v22.i34.7824.
- [4] H. M. H. Alam El-Din, A. S. Eldin, dan A. M. S. A. Hanora, "Bibliometric analysis of Egyptian publications on Hepatitis C virus from PubMed using data mining of an in-house developed database (HCVDBegy)," *Scientometrics*, vol. 108, no. 2, hal. 895–915, 2016, doi: 10.1007/s11192-016-2007-1.
- [5] T. H. S. Li, H. J. Chiu, dan P. H. Kuo, "Hepatitis C Virus Detection Model by Using Random Forest, Logistic-Regression and ABC Algorithm," *IEEE Access*, vol. 10, no. August, hal. 91045–91058, 2022, doi: 10.1109/ACCESS.2022.3202295.
- [6] H. Mamdouh Farghaly, M. Y. Shams, dan T. Abd El-Hafeez, "Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt," *Knowl. Inf. Syst.*, vol. 65, no. 6, hal. 2595–2617, 2023, doi: 10.1007/s10115-023-01851-4.
- [7] M. Yağanoğlu, "Hepatitis C virus data analysis and prediction using machine learning," *Data Knowl. Eng.*, vol. 142, 2022, doi: <https://doi.org/10.1016/j.datak.2022.102087>.
- [8] V. Viswanatha, A. C. Ramachandra, B. D. Parameshachari, S. V. Vardhini, dan N. Santhoshini, "Hepatitis C Disease Prediction Using Machine Learning Approach," *Int. Conf. Integr. Intell. Commun. Syst. ICIICS 2023*, hal. 1–6, 2023, doi: 10.1109/ICIICS59993.2023.10421118.
- [9] T. Simos, U. Georgopoulou, G. Thyphronitis, J. Koskinas, dan C. Papatoukas, "Analysis of protein interaction networks for the detection of candidate hepatitis B and C biomarkers," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 1, hal. 181–189, 2015, doi: 10.1109/JBHI.2014.2344732.
- [10] S. Vijayakumar, "A Hybrid QNN-Based Framework for Accurate Early Detection of HCV Liver Abnormalities from CT Scans Using Custom Transfer Learning and AI Edge Device," 2023 IEEE Reg. 10 Symp. TENSYP 2023, hal. 1–5, 2023, doi: 10.1109/TENSYP55890.2023.10223624.
- [11] A. Althemery et al., "Effectiveness comparison of patients with hepatitis C virus genotypes 1 or 4 therapies," *Clin. Epidemiol. Glob. Heal.*, vol. 31, no. May 2024, hal. 101880, 2025, doi: 10.1016/j.cegh.2024.101880.
- [12] L. Gutiérrez-Rojas, J. J. de la Gándara Martín, L. García Buey, J. I. Uriz Otano, Á. Mena, dan C. Roncero, "Patients with severe mental illness and hepatitis C virus infection benefit from new pangenotypic direct-acting antivirals: Results of a literature review," *Gastroenterol. Hepatol.*, vol. 46, no. 5, hal. 382–396, 2023, doi: 10.1016/j.gastrohep.2022.06.001.
- [13] D. Elreedy, A. F. Atiya, dan F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach. Learn.*, vol. 113, no. 7, hal. 4903–4923, 2024, doi: 10.1007/s10994-022-06296-4.
- [14] Y. Zhao dan C. Teng, "Classification of soil layers in Deep Cement Mixing using optimized random forest integrated with AB-SMOTE for imbalance data," *Comput. Geotech.*, vol. 179, no. December 2024, hal. 106976, 2025, doi: 10.1016/j.compgeo.2024.106976.
- [15] N. K. Pallicheruvu dan S. Gnanasekaran, "Energy Conversion and Management : X ANN-driven prediction of optimal machine learning models for engine performance in a dual-fuel mode powered by biogas and fish oil biodiesel," *Energy Convers. Manag.*, vol. 25, no. December 2024, hal. 100827, 2025, doi: 10.1016/j.ecmx.2024.100827.
- [16] J. Unjung dan M. R. Ningsih, "Optimized Handwriting-based Parkinson's Disease Classification Using Ensemble Modeling and VGG19 Feature Extraction," *Sci. J. Informatics*, vol. 10, no. 4, hal. 489–498, 2023, doi: 10.15294/sji.v10i4.47108.
- [17] G. A. Pradipta, R. Wardoyo, A. Musdholifah, dan I. N. H. Sanjaya, "Radius-SMOTE: A New Oversampling Technique of Minority Samples Based on Radius Distance for Learning from Imbalanced Data," *IEEE Access*, vol. 9, hal. 74763–74777, 2021, doi: 10.1109/ACCESS.2021.3080316.
- [18] R. Siringoringo dan I. K. Jaya, "Ensemble Learning Dengan Metode Smote Bagging Pada Klasifikasi Data Tidak Seimbang," vol. 3, no. 2, hal. 75–81, 2018.
- [19] J. Wei, Z. Lu, K. Qiu, P. Li, dan H. Sun, "Predicting drug risk level from adverse drug reactions using smote and machine learning approaches," *IEEE Access*, vol. 8, hal. 185761–185775, 2020, doi: 10.1109/ACCESS.2020.3029446.

- [20] H. T. Nguyen, L. T. Nguyen, A. D. Afanasiev, dan L. T. Pham, "Classification of Road Pavement Defects Based on Convolution Neural Network in Keras," *Autom. Control Comput. Sci.*, vol. 56, no. 1, hal. 17–25, 2022, doi: 10.3103/S0146411622010084.
- [21] N. Mohammad, A. M. Muad, R. Ahmad, dan M. Y. P. M. Yusof, "Accuracy of advanced deep learning with tensorflow and keras for classifying teeth developmental stages in digital panoramic imaging," *BMC Med. Imaging*, vol. 22, no. 1, hal. 1–13, 2022, doi: 10.1186/s12880-022-00794-6.
- [22] M. Si, T. J. Tarnoczi, B. M. Wiens, dan K. Du, "Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library-Keras: A Case Study on a Cogeneration Unit," *IEEE Access*, vol. 7, no. x, hal. 113463–113475, 2019, doi: 10.1109/ACCESS.2019.2930555.
- [23] T. Guillod, P. Papamanolis, dan J. W. Kolar, "Artificial neural network (ann) based fast and accurate inductor modeling and design," *IEEE Open J. Power Electron.*, vol. 1, no. August, hal. 284–299, 2020, doi: 10.1109/OJPEL.2020.3012777.
- [24] M. J. Nayeem, S. Rana, F. Alam, dan M. A. Rahman, "Prediction of Hepatitis Disease Using K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Multi-Layer Perceptron and Random Forest," 2021 *Int. Conf. Inf. Commun. Technol. Sustain. Dev. ICICT4SD 2021 - Proc.*, hal. 280–284, 2021, doi: 10.1109/ICICT4SD50815.2021.9397013.