Kufa Journal of Engineering
Vol. 17, No. 3, July 2026, P.P. 637-651
Article history: Received 19 October 2025, last revised 10 January 2026, accepted 10 January 2026

HEPATITIS C VIRUS STAGES CLASSIFICATION USING TRANSIT SEARCH ALGORITHM AND LONG SHORT TERM MEMORY

Zaineb M. Alhakeem[1], Heba Hakim[2], Hanadi Abbas Jaber[3], Suroor M. Dawood[4], Nuhad A. Malalla[5] , Zainab Ali Ashour Al-Tameemi[6] , Moamin Jawad Ahmed[7], Abbas Mohammed Hassan Lateef[8], and Khaledeh Mohammed Jawad[9]

[1] Department of Chemical and Petroleum Refining Engineering,Basrah University of Oil and Gas, Basrah, Iraq, Email:zainebalhakeem@buog.edu.iq.

[2] Computers Engineering Department, Engineering College, University of Basrah, Basrah, Iraq, Email:hiba.abdulzahrah@uobasrah.edu.iq.

[3] Computers Engineering Department, Engineering College, University of Basrah, Basrah, Iraq, Email:hanadi.jaber@uobasrah.edu.iq.

[4] Department of Chemical and Petroleum Refining Engineering,Basrah University of Oil and Gas, Basrah, Iraq, Email:suroor.moaid@buog.edu.iq.

[5] Department of Chemical and Petroleum Refining Engineering,Basrah University of Oil and Gas, Basrah, Iraq, email: dr.nuhad@buog.edu.iq.

[6] Computer Technology Department, Iraq University College, Basrah, Iraq, Email:zainabaltmimi0821@gmail.com.

[7] Computer Technology Department, Iraq University College, Basrah, Iraq, Email:alhydrymwmn@gmail.com.

[8] Computer Technology Department, Iraq University College, Basrah, Iraq, Email:abbasmohammed90.h@gmail.com.

[9] Computer Technology Department, Iraq University College, Basrah, Iraq, Email:khaledeh.mj@gmail.com.

https://doi.org/10.30572/2018/KJE/170336

ABSTRACT

Hepatitis C Virus (HCV) is a disease that infects the liver with multiple stages that spread through blood, requiring blood tests and body symptoms for diagnosis. The diagnosis should decide which stage the patient reaches. This work suggests a hybrid classification method based on optimization and machine learning techniques. A Long Short-Term Memory Neural

Network (LSTM) is used as a classifier to identify the stages of the disease, based on the four stages, using 28 features comprising body symptoms and blood tests. To find the optimal number of hidden cells in the hidden layers of LSTM, the Transit Search Algorithm (TSA) is used, TSA identifies the best integer value for the number of hidden cells and selects the best features combined with this number of cells that will give the highest accuracy of classification. The preprocessing step is required to enhance the quality of the dataset and resizing it. In multiclass classification, a large dataset is essential for the network to learn effectively through all classes. To achieve this, Synthetic Minority Oversampling Techniques (SMOTE) is used to generate similar data that will increase the size of the dataset. The proposed TSA-LSTM method achieves high performance, with classification accuracy exceeding 99% outperforming previous works.

KEYWORDS

Hepatitis C Virus, Machine Learning, Disease Prediction, Transit Search Algorithm, Long Short-Term Memory.

1. INTRODUCTION

The medical applications of disease diagnosis are very useful field in medical field that will shorten the diagnosis time and help the medical professions to make there dissensions more reliable based on Machine Learning (ML) Models that trained on hundreds of different cases. Some studies are used ML to diagnose the autism, heart diseases , bone fractions or other diagnosis models (Alhakeem et al., 2023, 2025; Alzubaydi & Alyasseri, 2025; Hashim & Mazinani, 2025; Mansour et al., 2025).

Hepatitis C (HC) is an infection caused by the Hepatitis C Virus (HCV), which attacks the liver and spreads by any amount of blood. This infection can lead to liver inflammation, can result in both acute and chronic hepatitis. This disease may last either for a few weeks or for a lifetime (Chigbu et al., 2019). According to statistics made by the World Health Organization (WHO) following the initial infection, almost 80% of patients show no symptoms appears on them (Hepatitis C, n.d.). As there is no vaccine for HC, early diagnosis is needed to save patients and people who are in contact with them.

The contribution of this work lies in identifying the necessary body features and tests that should be available to classify liver disease into four different stages. The classification accuracy is about 98% for both classification with and without optimization.

2. LITRETURE REVIEW

In this section a previous studies that focuses on the HC disease are listed and discussed to be compared to this work. Many types of ML are used as will be shown for many reasons (diagnosis or classification).

Kurniawan et al. proposed a method to classify the patients as infected with either Hepatitis B or C according to their medical records that are collected in the hospital. They used K-Mean Clustering for classification and tested their method using different sizes of training and testing groups, they found that 70:30 is the best ratio that will give an 84% classification accuracy. This study does just separate the types of hepatitis and it does not take into consideration multiclass datasets or compare its result with other works (Kurniawan & Rustam, 2019).

Some studies used very special blood tests that could not be done by a non-specialist laboratory such as in(Haga et al., 2020), and the research group classified the disease according to genome DNA using a Support Vector Machine that fed 81 features. In (Sohail et al., 2018) the authors used multiple classifiers and feature selection method for comparison, using a dataset of infected blood serum and comparing it with healthy blood. These works are so special that patients and doctors cannot be familiar with it.

Ahammed et al. used K Nearest Neighbor (KNN) algorithm to classify a collected dataset from Egyptian patients (Ahammed et al., 2020). This dataset has five classes with 29 features . They used Synthetic Minority Oversampling Techniques (SMOTE) to increase the balancing of the dataset and increase the size five times. This process improved the classification accuracy from about 25% to 94%. Additionally, they performed feature selection to find out the most relevant features for best classification accuracy.

Mamdouh et al. used a Random Forest (RF) algorithm with Sequential Forward Selection (SFS) to classify a group of Egyptian patients into HCV patients or not and compared their findings with other methods, they could achieve about 95% using four features for binary classes (Mamdouh Farghaly et al., 2023). and Eden et al.(Edeh et al., 2022) proposed an Ensemble Learner (EL) consisting of multiple classifiers, which could classify multiple classes with high accuracy that reaches 95%. Ali et al. in (Ali et al., 2023) used SFS with Machine Learning (ML) to classify binary classes of a 13 features dataset, in this work the dataset is balanced and increased using SMOTE.

Bhargav et al. and Yakubu et al. used a binary classes dataset that contains HCV or die to predict whether there is a chance that the patient will die or simply has HCV. Although their method achieved high accuracy of classification, the overall application of such classification is not very practical to fasten the diagnosis of the disease(Bhargav et al., 2018) (Yakubu & Nsang, 2022).

In (Wang et al., 2023), Wang and the research team used Adaptive Boost (AdaBoost) with Principal Component Analysis (PCA) to classify three categories: health, suspected, and HCV. Neural Networks (NNs) are widely used for classification because it is better to select features without any need for manual selection. Syafa'ah et al. compared multiple types of machines for the classification of HCV and found that NNs were the best choice to do the task (Syafa'ah et al., 2021). Butt et al. also used NN for the multiclassification of HCV, manually select 19 features out of 29. However, they did not mention any folding technique to change the order of the observation (Butt et al., 2021).

3. METHODOLOGY

This section represents the overall system proposed in this work, as shown in Fig. 1. The dataset acquisition, feature selection, and feature classification are explained in detail in the following subsections.

3.1. Dataset

HCV datasets are available from laboratories and hospitals because of the need for blood tests. Early symptoms are very common and do not give a sharp decision of prediction. The used

dataset in this work is the Egyptian dataset, which has five classes and 1385 samples, available in the UCI Repository (Kamal et al., 2019). Table 1 lists the features and their descriptions. Eleven features are the common body symptoms such as (Age, Gender, etc) and the other seventeen are related to the blood tests. This data has no missing values, so no need for a cleaning process before applying the method. It is divided into two groups with the ratios of Training set =80% and Testing set =20%.

Table 1. Features of the HCV Dataset

Feature	Description
Age	31-62 years
Gender	Male=1/Female=2
BMI(Body Mass Index)	22-35
Fever	Absent=1, Present=2
Nausea/Vomiting	Absent=1, Present=2
Headache	Absent=1, Present=2
Diarrhea	Absent=1, Present=2
Fatigue	Absent=1, Present=2
Bone ache	Absent=1, Present=2
Jaundice	Absent=1, Present=2
Epigastria pain	Absent=1, Present=2
WBC(White Blood Cells)	2991-12101
RBC(Red Blood Cells)	3816422-5018451
HGB(Hemoglobin)	2-20
Plat(Platelet)	93013-226464
AST1(1 week)	39-128
ALT1(1 week)	39-128
ALT4(4 weeks)	39-128
ALT12(12 weeks)	39-128
ALT24(24 weeks)	39-128
ALT36(36 weeks)	39-128
ALT48(48 weeks)	39-128
RNA Base	0-1201086
RNA 4	0-1201715
RNA 12	0-3731527
RNA EOT	0-808450
RNA EF (Elongation Factor)	0-808450
Baseline Histological Grading	1-16
Baseline Histological	No Fibrosis=0, Portal Fibrosis=1,Few Septa=2,Many Septa=3, Cirrhosis=4

3.2. Preprocessing

The data is pre-processed before being used for the final test. First, the data is resized using the Synthetic Minority Oversampling Technique (SMOTE). This technique will simulate the data and generate a similar and bigger dataset, in this work, the dataset is resized form1386 the sample to 13032.Next, the normalization process is applied to data using the zero-score normalization method.

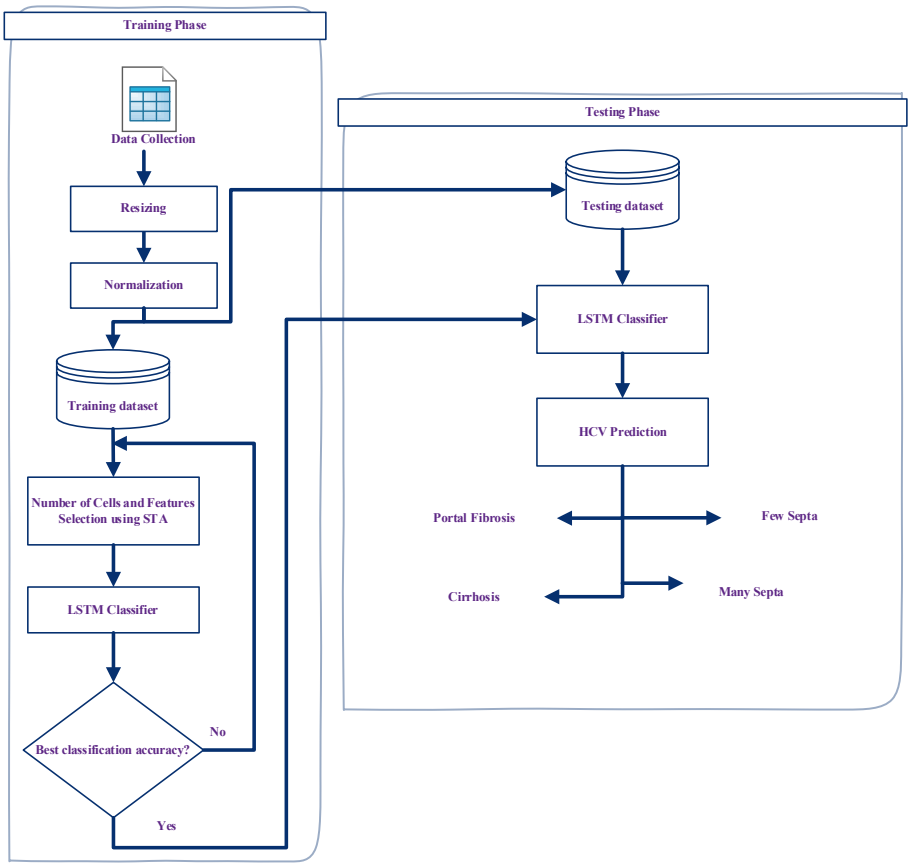

Fig. 1. The Overall proposed system

3.3. Features Selection

Features selection is performed using the Transit Search Algorithm (TSA). Both the number of hidden cells and the features are considered as features in the optimization step. The number of hidden cells is represented by integer values, while the features selection is represented by binary values.

This algorithm is an astrophysics-inspired metaheuristic algorithm that is proposed based on the xoplanet exploration method. It is suitable for both high and low-dimension problems. The algorithm has only two parameters: the number of stars and Signal to Noise Ratio (S/N). It has high results with low complexity compared to the common algorithms (Mirrashid & Naderpour, 2022). The TSA consists of five stages the galaxy, the star phase, the transit phase, the planet phase, the neighbour phase, and finally the exploitation phase.

Firstly, in the galaxy and star phase, TSA randomly selects the location of the galaxy centre and identifies the habitable zones of the galaxy. After selecting a region, a star is selected to find a stellar system. By the end of these phases, there will be several stars to search by algorithm. These phases are done once for all iterations to find out the best situation to execute the subsequent three phases. In the second phase, the amount of light received from the star is measured to detect any possible reduction in the received light signals. In the transit phase, the

amount of light that received from a star should be remeasured to find out the type of star. The star with higher brightness is equal to the higher fitness and the probability of the transit is calculated in this phase to be used in the next phase.

The initial location of the detected plant is determined in the plant phase. The SN parameter is one of the important parameters in this algorithm which is used to reduce the noise impact. The approximation situation of the plant is determined in this phase based on the number of the received signals that indicates the plant's location. In TSA, the best plant for each star is determined then after this step the neighbour phase will be started. If no transit is detected for a star, the nearby plants that have already identified for the same star will be studied. Finally, in the exploitation phase, the best plant that could host life is identified which represents the global solution. Fig. 2 shows the flowchart of the TSA in details.

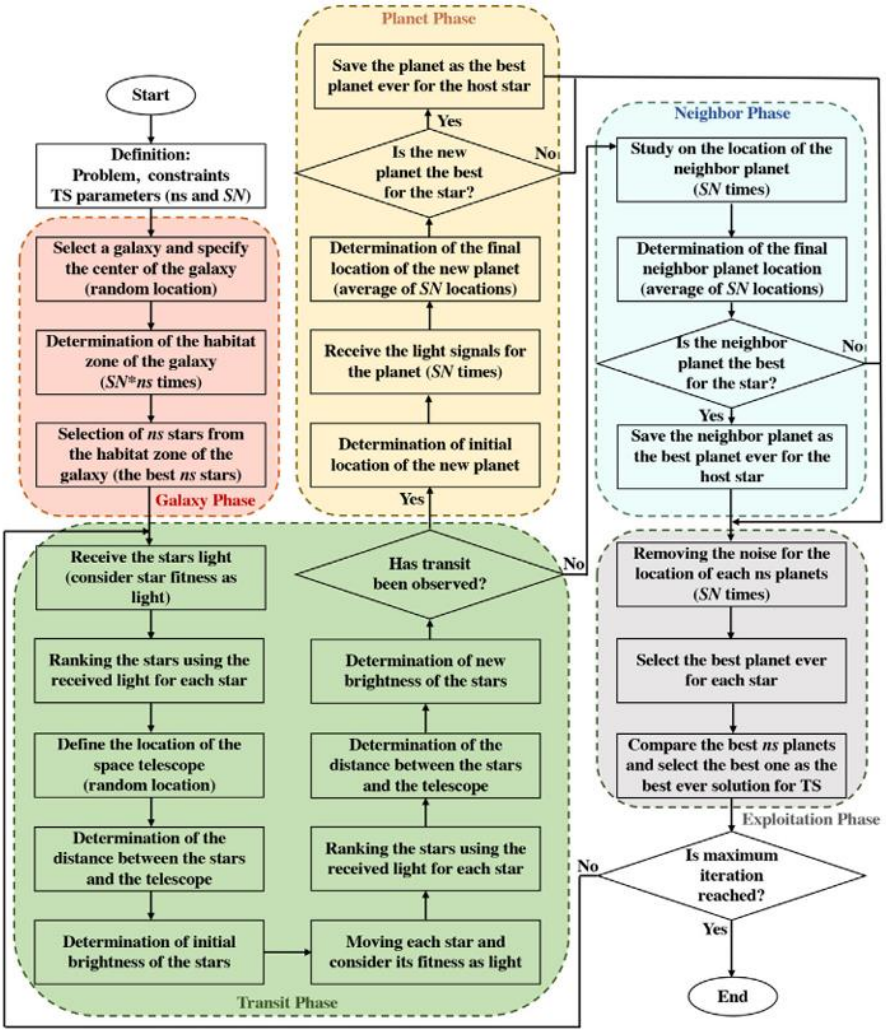

Fig. 2. The flowchart of TSA (Mirrashid & Naderpour, 2022).

3.4. Long Short-Term Memory

One of the recurrent networks is the long short-term memory that used to solve the problem of back-flow. Four vectors are introduced in this network which they are: Input Activation, Output

Activation, Forget Activation, and the last vector is the Candidate. Fig. 3 shows the structure of a single cell of this LSTM, where the input sequences are h_k, x_k *and* $c_k$are the cell output vector, the input sequence, and the cell state vector to the network cell, respectively. The vectors are applied to the activation function (sigma) and the function (hyperbolic tangent (tanh)) will compute the states of the cellsc_k (Hochreiter & Schmidhuber, 1997).

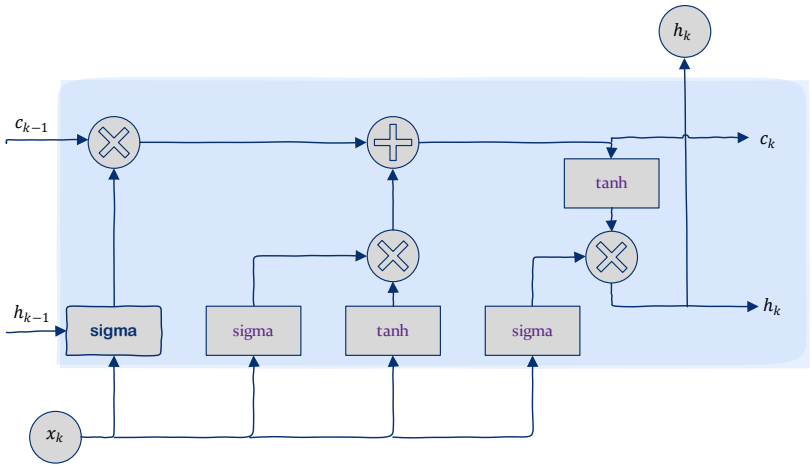

Fig. 3. The hidden cell of the network

4. PROPOSED METHOD

In this work, a combination of LSTM network and TSA algorithm is used for features selection and classification. LSTM is employed to extract features from the dataset. Although the dataset includes features of body and blood tests, not all of these features are very important for this task, so a features selection method is necessary. LSTM used in this work has two hidden layers and five other layers. Besides feature selection, an algorithm is used to select the number of cells in each layer. Fig. 4 illustrates the structure of the NN, which consists of seven layers: input layer, two hidden layers, a fully connected layer, Softmax layer, and finally the output layer. The ADAM optimizer is used with 12 samples as a minimum batch size.

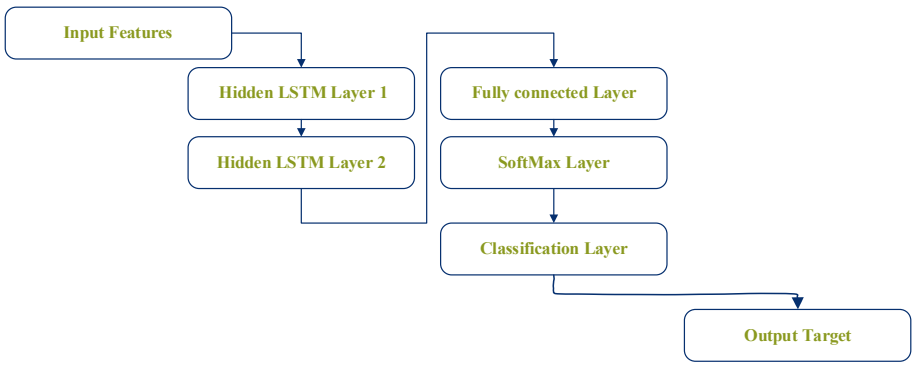

Fig. 4. The proposed LSTM structure

To achieve the best accuracy, TSA is used to both optimize feature selection and number of cells in the hidden layers in the same time. The parameters for the selection algorithm are listed in Table 2. This algorithm has few parameters, making it easy to choose appropriate values.

With three stars and 10 SN, there are a total of 30 search agents. The algorithm is repeated 100 times to find the best solution that minimizes the classification error of LSTM, using a combination of integer and binary values to represent the number of cells and selected features, respectively.

Table 2. TSA Parameters

Parameter	Value
Number of stars	3
Signal-to-Noise Ratio	10
Max iterations	100

After the selection step is completed, the final parameters of the LSTM are listed in the Table 3. This network is then used in the classification of the liver disease stages.

Table 3. Classifier Parameters

Parameter	Value
Number of layers	2
Number of hidden cells layer 1	500
Number of hidden cells layer 2	500
Number of Iteration	100
Minibatch	12
Optimizer	ADAM

5. RESULTS AND DISCUSSION

The following subsections will show the impact of various factors, including the number of classes in the dataset, the size of the dataset, the number of features, and the type of optimization algorithm and classifier.

5.1. Case 1: Binary Class Dataset Classification

First, we test the dataset with only two classes: the presence of liver disease or the absence of the disease. For this test, the dataset is converted to binary class with 28 features. LSTM classifier with two hidden layers and different hidden cells are used. Table 4 list the accuracy of different numbers of hidden cells, which were chosen manually. The results are almost the same but not very satisfying, indicating that more processing and careful choosing of parameters are needed to improve the accuracy of classifying.

Table 4. Accuracy of LSTM

Number of hidden cells	Accuracy
10	67.9
25	66.8
50	68.2
100	66.4
200	69.7
300	69.3
400	66.8
500	64.6

Instead of manually choosing the number of hidden cells and the features for NN, an optimization algorithm is used for this task. The Whale Optimization Algorithm is employed to

simultaneously find the optimal number of hidden cells and the best relative features, aiming to get high accuracy. WOA identified that 11 features are the relevant features with 436 hidden cells for the first hidden layer and 248 hidden cells for the second layer of LSTM.

Gated Recurrent Unit (GRU) is tested for comparison with WOA, using 181 and 447 cells for each hidden layer with 10 features only.

Fifty runs with folding the data in each one are conducted to evaluate the performance across different sets. Table 5 lists the minimum, maximum, and average performance of both WOA-LSTM and WOA-GRU. The accuracy of both combinations increased by about 4%. Although Despite the increasing of the accuracy, the difference between the minimum and maximum values is about 10% for both methods.

Table 5. The results of WOA-NN in the binary class

Classifier	Average Accuracy	Maximum Accuracy	Minimum Accuracy
WOA-LSTM	73.8	79.4	68.9
WOA-GRU	74.11	79.4	68.2

Although the results show an increase in accuracy, the binary class dataset does not give a good decision in diagnosis for the stages of liver disease. Therefore, a multi-class approach will be used for the rest of this work.

5.2. Case 2: Multiclass Small Dataset Classification

In disease prediction, the choice between multiclass and binary class depends on the type of the disease. HCV is a liver disease that has different categories, not just infected or not. Knowing the types or stages of liver disease is better and more useful than just classifying the presence or absence of the disease.

The Egyptian data set has four different stages of HSV, which makes it suitable for testing the classifier using small-size dataset with 28 features. Table 6 shows the results of using LSTM with 10 and 100 hidden cells in two hidden layers. The performance is very low and unacceptable with this dataset. The small-size dataset is insufficient to accurately recognize the classes, so increasing the dataset size should be solve this issue.

Table 6. Results of LSTM using multiclass dataset

Classifier	Average Accuracy	Maximum Accuracy	Minimum Accuracy
10 hidden cells	25.3	28.7	21.5
100 hidden cells	25.3	28.3	22.3

5.3. Case 3: Multiclass Big Dataset Classification

In this case, two classifiers LSTM and GRU are tested, each with a single layer containing 100 hidden cells with all 28 features. The test is repeated 50 times, with the dataset being folded each time to find out the generality status of the classifier. Table 7 listed the average accuracy, as well as the minimum and maximum accuracy values of both LSTM and GRU. It can be seen

that LSTM and GRU have almost the same average accuracy but LSTM has a higher maximum value and a lower minimum value. So, LSTM will be chosen in this work.

Table 7. Comparison between LSTM and GRU

Classifier	Average Accuracy	Maximum Accuracy	Minimum Accuracy
GRU	98.217	98.81	97.27
LSTM	98.4	99.23	96.54

TSA was tested twice: initially using only one iteration with 500 hidden cells and 27 features which improve performance. The feature "Bone ache" is not a very relevant feature so it excluded form classification. In the second test, TSA used 100 iterations instead of one algorithm, finding that 26 features were the relevant features, with Jaundice and Plat (Platelet) being unnecessary for the classification. Fig. 5 shows the error of the classification during 100 cycles, illustrating that the optimal solution was found after 30 cycles.

The average accuracy of TSA-LSTM during 50 runs using both parameters of one cycle TSA and 100 cycles, is shown in Table 8. The value of accuracy during 50 runs ranged from a minimum of 98.5% to a maximum of 99.5%, with the variance of less 1%. That indicates that LSTM with the parameters selected by TSA, is general and can give the same result using any combination of datasets.

Table 9 compares previous works in the same field of Liver disease classification. Some works have achieved very good performance with binary classes, reaching a maximum accuracy of about 96% as seen in (Yakubu & Nsang, 2022).

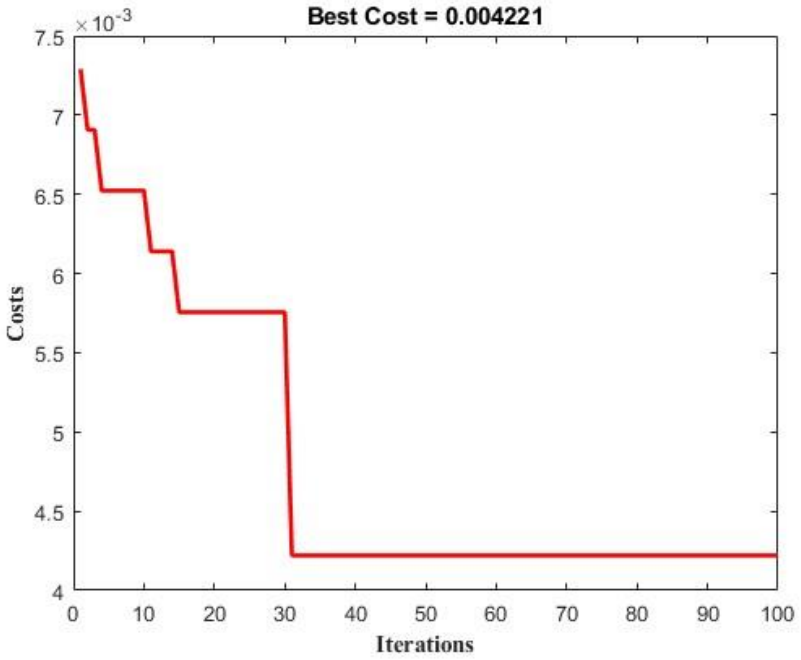

Fig. 5. The cost of the TSA during 100 cycles

However, binary classifying can only give us the indicator about there is infection or not without predicting of stages of the infection ,unlike multiclass classification. Other works that deal with multiclass classification using small-size datasets have also a very good performance, reaching the performance more than 95%. In this work, our goal is to achieve a high-performance classifier regardless of the number of features. Large-size datasets are expected to give a high accuracy and enable the classifier to compete with other proposed methods. The high-performance classifier will be a reliable one of making diagnosis applications or predictors in future work.

Table 8. TSA-LSTM performance using a Large dataset

Classifier	No. of TSA Iterations	Number of Features	Hidden cells 1	Hidden Cells 2	Average Accuracy
TSA-LSTM	1	27	500	500	98.8734
TSA_LSTM	100	26	500	28	99

Table 9. Comparison with previous works

Reference	Algorithm	Number of Features	Number of classes	Accuracy%
(Bhargav et al., 2018)	LR	20	2	87.17
(Yakubu & Nsang, 2022)	DT	14	2	96
(Mamdouh Farghaly et al., 2023)	RF SFS	4	2	94.99
(Ahammed et al., 2020)	KNN	29	5	94.40
(Edeh et al., 2022)	EL	13	5	95.59
(Butt et al., 2021)	NN	18	5	94.44
	GRU	28	4	98.2
	LSTM	28	4	98.4
	WOA-LSTM	11	4	73.8
	WOA-GRU	10	4	74.1
Proposed Method	TSA-LSTM	26	4	99

6. CONCLUSIONS

This work proposed a hybrid model consisting of an optimization algorithm and a neural network for feature selection and classification, respectively. Long Short-Term Memory (LSTM) was used for classifying the multiclass dataset of HCV, while the Transit Search Algorithm (TSA) was used to optimize the selection of both the number of hidden cells in the LSTM's hidden layers and the relevant features to get the best performance.

Form previous studies the binary classification is easier than multiclass classification because of the relationship among the classes and features, in this work LSTM is used to do the classification process. We conclude that small datasets can not be enough to give high performance classification even when the model trained enough. When we increase the size of the dataset using SMOTE technique the accuracy enhanced to reach more than 95%. TSA is used to decrease the number of features and eliminate the unrelated features, beside that increase

the complexity of the model not always give the wanted results. When we decrease the number of hidden cells in the second hidden layer the performance increased to reach about 99% which is very promising value in multiclass classification. These results are good enough to design and implement a reliable diagnostic application in future work.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

7. REFERENCES

Ahammed, K., Satu, Md. S., Khan, Md. I. and Whaiduzzaman, M. (2020) Predicting Infectious State of Hepatitis C Virus Affected Patient's Applying Machine Learning Methods. (2020) IEEE Region 10 Symposium (TENSYMP), 1371-1374. https://doi.org/10.1109/TENSYMP50017.2020.9230464.

Alhakeem, Z. M., Dawood, S. M., Homod, R. Z. and Mohammed, H. I. (2025) Design and Execution of an Application for Early-Stage Stroke Prediction Using Ensemble Learners. Proceedings of (2025) IEEE International Symposium on Future Telecommunication Technologies, SOFTT 2025, 25-30. https://doi.org/10.1109/SOFTT67007.2025.11213397.

Alhakeem, Z. M., Hakim, H., Hasan, O. A., Laghari, A. A., Jumani, A. K., Jasm, M. N., Alhakeem, Z. M., Hakim, H., Hasan, O. A., Laghari, A. A., Jumani, A. K. and Jasm, M. N. (2023) 'Prediction of diabetic patients in Iraq using binary dragonfly algorithm with long-short term memory neural network', AIMS Electronics and Electrical Engineering (2023) 3:217, 7(3), 217-230, available: https://doi.org/10.3934/ELECTRENG.2023013.

Ali, A. M., Hassan, M. R., Aburub, F., Alauthman, M., Aldweesh, A., Al-Qerem, A., Jebreen, I. and Nabot, A. (2023) 'Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection', Machines, 11(3), 391, available: https://doi.org/10.3390/machines11030391.

Alzubaydi, H. and Alyasseri, Z. (2025) 'ENHANCED EEG SIGNAL CLASSIFICATION FOR AUTISM SPECTRUM DISORDER USING RIEMANNIAN GEOMETRY AND ENSEMBLE LEARNING TECHNIQUES', Kufa Journal of Engineering, 16(3), 176-198, available: https://doi.org/10.30572/2018/KJE/160312.

Bhargav, K. S., Thota, D., Kumari, T. D. and Vikas, B. (2018) 'Application of machine learning classification algorithms on hepatitis dataset', International Journal of Applied Engineering Research, 13, 12732, --12737.

Butt, M. B., Alfayad, M., Saqib, S., Khan, M. A., Ahmad, M., Khan, M. A. and Elmitwally, N. S. (2021) 'Diagnosing the Stage of Hepatitis C Using Machine Learning', Journal of Healthcare Engineering, 2021, 1-8, available: https://doi.org/10.1155/2021/8062410.

Chigbu, D. I., Loonawat, R., Sehgal, M., Patel, D. and Jain, P. (2019) 'Hepatitis C Virus Infection: Host-Virus Interaction and Mechanisms of Viral Persistence', Cells, 8(4), 376, available: https://doi.org/10.3390/CELLS8040376.

Edeh, M. O., Dalal, S., Dhaou, I. Ben, Agubosim, C. C., Umoke, C. C., Richard-Nnabu, N. E. and Dahiya, N. (2022) Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.892371.

Haga, H., Sato, H., Koseki, A., Saito, T., Okumoto, K., Hoshikawa, K., Katsumi, T., Mizuno, K., Nishina, T. and Ueno, Y. (2020) A machine learning-based treatment prediction model using whole genome variants of hepatitis C virus. PLOS ONE, 15(11), e0242028. https://doi.org/10.1371/journal.pone.0242028.

Hashim, A. A. and Mazinani, M. (2025) 'DETECTION OF KERATOCONUS DISEASE DEPENDING ON CORNEAL TOPOGRAPHY USING DEEP LEARNING', Kufa Journal of Engineering, 16(1), 463-478, available: https://doi.org/10.30572/2018/KJE/160125.

Hepatitis C. (n.d.) Retrieved January 5, 2026, from https://www.who.int/news-room/fact-sheets/detail/hepatitis-c.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', Neural Computation, 9(8), 1735-1780, available: https://doi.org/10.1162/neco.1997.9.8.1735.

Kamal, S., ElEleimy, M., Hegazy, D. and Nasr, M. (2019) Hepatitis C Virus (HCV) for Egyptian patients. UCI Machine Learning Repository.

Kurniawan, G. and Rustam, Z. (2019) Enhancement of hepatitis virus outcome predictions with application of K-means clustering. 020044. https://doi.org/10.1063/1.5132471.

Mamdouh Farghaly, H., Shams, M. Y. and Abd El-Hafeez, T. (2023) 'Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt', Knowledge and Information Systems, 65(6), 2595-2617, available: https://doi.org/10.1007/s10115-023-01851-4.

Mansour, H. S., Valizadeh, M., Abdulaal, A. H. and Amirani, M. C. (2025) 'A Novel Deep 2D-CNN Model for ECG-Based Arrhythmia Diagnosis with Selective Attention Mechanism and CWT Integration', Kufa Journal of Engineering, 16(2), 423-444, available: https://doi.org/10.30572/2018/KJE/160225.

Mirrashid, M. and Naderpour, H. (2022) 'Transit search: An optimization algorithm based on exoplanet exploration', Results in Control and Optimization, 7, 100127, available: https://doi.org/10.1016/j.rico.2022.100127.

Sohail, A., Khan, S., Ullah, R., Qureshi, S. A., Bilal, M. and Khan, A. (2018) Analysis of hepatitis C infection using Raman spectroscopy and proximity based classification in the transformed domain. Biomedical Optics Express, 9(5) (2041) https://doi.org/10.1364/BOE.9.002041.

Syafa'ah, L., Zulfatman, Z., Pakaya, I. and Lestandy, M. (2021) 'Comparison of Machine Learning Classification Methods in Hepatitis C Virus', Jurnal Online Informatika, 6(1), 73, available: https://doi.org/10.15575/join.v6i1.719.

Wang, Y., Yin, B. and Zhu, Q. (2023) Application of Machine Learning Algorithms in Predicting Hepatitis C. Proceedings of the (2023) 4th International Symposium on Artificial Intelligence for Medicine Science, 359-365. https://doi.org/10.1145/3644116.3644176.

Yakubu, Y., & Nsang, A. S. (2022). Predicting Hepatitis Using Decision Trees. *IRE Journals*