

An Efficient Method For Classification of Different Types of Hepatitis Virus Using Genomic Signal Processing and Machine Learning

**Love Fadia, Vatsal Shah, Mohammad Hassanzadeh,
Jonathan Wu and Majid Ahmadi**

Department of Electrical and Computer Engineering

University of Windsor

Windsor, Ontario, Canada

fadial@uwindsor.ca, shah7r1@uwindsor.ca,

mhassan@uwindsor.ca, jwu@uwindsor.ca

and M.ahmadi@uwindsor.ca

Abstract

This paper describes an efficient method for classifying the four main hepatitis virus types Hepatitis-B, Hepatitis-C, Hepatitis-D, and Hepatitis-E using Genomic Signal Processing and Machine Learning. First, we gather the deoxyribonucleic acid sequences for the various strains of the Hepatitis virus. Next, we convert these sequences from characters to numbers using a variety of coding algorithms. Subsequently, We employ some well-known signal processing methodologies along with some modified versions of signal processing developed by ourselves to extract the characteristics from the transformed sequences and then we use Singular Value Decomposition for dimensionality reduction. Ultimately, two machine learning models Decision Tree and Light Gradient Boosting Machine are trained for classification. Our approach gets an accuracy of 99% with the combination of normalization of atomic number with a customized Haar Wavelet.

Keywords

Genomic signal processing, Fast Fourier Transform, Haar Wavelet Transform, Coiflet Wavelet Transform, Machine Learning algorithms.

1. Introduction

Hepatitis is an inflammatory condition of the liver usually brought on by viral infections. The distinctiveness of hepatitis A, B, C, D, and E lies in the many viral strains that produce the illness. There are several ways in which they can spread, including through blood-to-blood contact or the ingestion of tainted food or drink (Tu. Et al. 2017). Proper categorization and comprehension of each kind are essential for the effectiveness of treatment regimens, prophylactic actions, and public health administration. This can be accomplished using Genomic signal processing (GSP) and machine learning which are critical in better understanding the patterns in hepatitis virus because they allow more exact categorization and analysis of the many viral strains. Different genetic patterns and mutations associated with different hepatitis viruses can be discovered via this technique, as it converts biological data such as DNA and RNA sequences into digital signals. This allows for more exact identification of certain virus types and subgroups, which is required for the development of effective treatments and vaccines. (Dougherty et al. 2009). Therefore here we proposed an efficient way of classification of Hepatitis virus based on GSP.

1.1 Objectives

- To get important features from the Hepatitis Virus so that the classification task becomes easier.
- To find out whether modifying certain signal processing algorithms is it possible to get more accurate results or not.
- To investigate that by changing in certain coding techniques can we get more precise numbers that can be used by signal processing algorithms to generate features.

2. Literature Review

Deep learning-based unsupervised clustering of DNA sequences was demonstrated by Arias et al. (2022), who utilized the Frequency Chaos Game Representation (FCGR) technique. This approach involves converting the FCGR matrix into an image format, which was then analyzed using an Artificial Neural Network to effectively classify Hepatitis, Influenza, and Dengue viruses. Additionally, Lilhore et al. (2023) developed a hybrid predictive model that assists experts in classifying the Hepatitis C virus (HCV), enabling the identification of various HCV genotypes, as further explored by Fahmy et al. (2024). Despite its capability to manage a 10-class classification, this model encountered limitations due to class imbalance, with uneven numbers of DNA sequences for each genotype. A clinical data-based approach to differentiate the Hepatitis B virus (HBV) from normal cases was proposed by Breckenridge et al. (2019). In this method, the model is trained on specific parameters derived from patient tests. Moreover, Mamdouh et al. (2022) presented a method for classifying the HCV virus using 12 clinical dataset features, tested both with and without feature selection, employing machine learning classifiers such as Naïve Bayes (NB), K-nearest neighbor (KNN), Random Forest (RF), and Logistic Regression (LR). Finally, Nilashi et al. (2019) discussed the feasibility of classifying Hepatitis virus using fuzzy networks, noting that the limited amount of available data was a significant constraint.

3. Methodology

We have collected data on Hepatitis B, C, D, and E from the National Centre of Biotechnology Institute (NCBI) website which is an open-source tool where different organization all over the world publishes the DNA sequence that has been generated. It consists of the sequences ranging from Dengue to COVID-19. We gathered 1000 sequences of each subtype of the Hepatitis virus having a length of 1000. After that, it was converted into numbers using three techniques namely Electron-Ion-Potential(EIIP), Normalization of Atomic Number(NAN), and Complex coding. Once we get different numeric sequences we apply four signal processing techniques which are, Discrete Sine Transform (DST), Haar Wavelet Transform, Customized Haar Wavelet Transform, and Fast Fourier Transform (FFT) to extract features. Then we applied Singular Value Decomposition (SVD) which helps to reduce the dimension of the long features while retaining as much information as it can and after that, the data is split into two parts. 70% of the data is used for training the two machine learning models and the rest is used for testing purposes. The Figure 1. illustrates the same

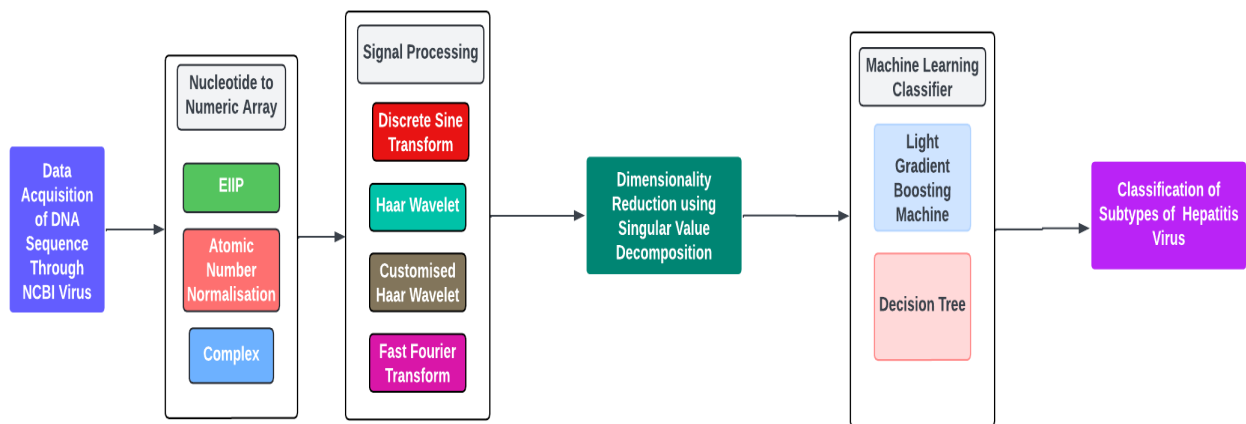


Figure 1. Flowchart of Proposed Methodology

3.1 Coding Technique

EIIP Coding:

The DNA sequence can be converted into floating points using the Electron-Ion potential value assigned to the nucleotide of the sequence (Sahu.et.al,2011) which are:

Table 1. EIIP Coding Technique

Nucleotide	Symbol	EIIP Value
Adenine	A	0.1260
Guanine	G	0.0806
Thymine	T	0.1335
Cytosine	C	0.1340

Normalization of Atomic Number:

Another approach is changing the DNA sequence into floating points based on the atomic number of each nucleotide present in the sequence once we assigned numbers then we normalized it. The number that each nucleotide assigned is:

Table 2. Normalization of atomic numbers

Nucleotide	Symbol	Normalized Value
Adenine	A	0.257
Guanine	G	0.287
Thymine	T	0.243
Cytosine	C	0.213

Complex Coding:

Based on the complementary property of DNA sequence one can also apply complex conjugate numbers to the nucleotide(Akhtar. et.al, 2008).

Table 3. Complex Coding

Nucleotide	Symbol		Complex Number
Adenine	A		1+i
Guanine	G		-1+i
Thymine	T		-1-i
Cytosine	C		1-i

3.2 Signal Processing

Discrete Sine Transform

The DST represents the sinusoidal terms of the Fourier transform of varying frequency. It has good energy compaction and it has 4 different versions, but we utilized version II which is shown in equation 1 as follows (Britanak.et.al, 2007).

$$X_k = \sum_{n=1}^N x_n * \sin \left[\frac{\pi}{N+1} nk \right] \quad (1)$$

Where,

- X_k the transformed coefficient at index k.
- x_n : the n-th element of the input sequence.
- N: the length of the input sequence.
- k: the index ranging from 1 to N.

Haar Wavelet Transform

Haar wavelet transform is a function of rescaled square shapes. It is one of the simplest wavelet transforms. Here we implemented this transformation at level 4. The transformation can be represented as: where the first row represents scaling coefficients and from the second to the fourth row it represents various levels' coefficients (Patrick.et.al,2019).

$$H_4 = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{pmatrix} \quad (2)$$

Customized Haar Wavelet Transform

Here we customized the coefficients of the Haar wavelet transform to achieve better accuracy. The customization is represented in the form of a matrix as: This matrix can be applied to the signal vector x to produce the wavelet coefficients figure 2 below illustrates the workflow of customized haar wavelet .

$$W = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \quad (3)$$

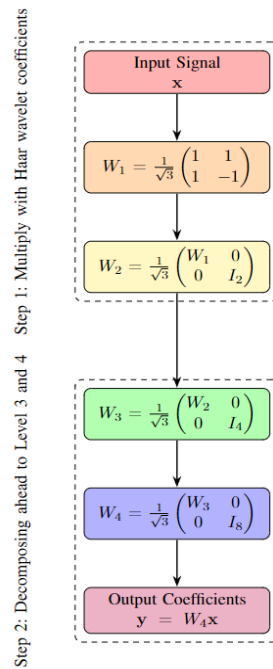


Figure 2. Workflow for Customized Haar Wavelet

Fast Fourier Transform

The FFT is the algorithm that computes the discrete Fourier transform of a signal at an optimal rate. The equation for the FFT (Hassanzadeh.et.al, 2022) is:

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] \cdot e^{-j\left(\frac{2\pi}{N}\right)kn} \quad (4)$$

Where,

- $X[k]$: the k-th component of the frequency domain.
- $x[n]$: the n-th element of the input time domain sequence.
- N : the total number of points in the sequence.
- \sqrt{N} : the normalization factor that makes the transform unitary, preserving the energy of the signal.
- $(2\pi/N)kn$: the phase factor that correlates the time index n and frequency index k .

3.3 Dimensionality Reduction

SVD is a mathematical technique used to factorize a matrix into three distinct matrices: U , Σ , and V^T . Given a matrix A , SVD expresses it as $A = U\Sigma V^T$ where U and V are orthogonal matrices representing the left and right singular vectors, respectively, and Σ is a diagonal matrix containing the singular values of A (Klema. et.al, 1980). These singular values represent the magnitude of the directions defined by the singular vectors. SVD is widely used in data compression, noise reduction, and dimensionality reduction, as it allows for the approximation of a matrix by retaining only the largest singular values and their corresponding vectors, effectively capturing the most significant features of the data (Schmidt.et.al,2020). Here we have taken the U matrix as input for the classifier model, taking the first 5 rows of the U matrix. The reason for choosing the U matrix is because it has eigenvectors rather than values and it is found that more significant variations are stored in the U matrix.

3.4 Machine Learning Model

Light Gradient Boosting Machine

Using the idea of boosting, an ensemble learning technique called a Light Gradient Boosting Machine (LGBM) is employed. The LGBM model operates on a similar principle to other boosting algorithms. The model successively builds an ensemble of decision trees while ensuring that each tree gains knowledge from the errors of the trees built in earlier iterations. In the end, the weighted total of all the decision trees that were built is used to create the final model (Osman. et.al, 2021). Moreover, unlike traditional level-wise tree development, which increases the tree's levels one at a time, LGBM uses a leaf-wise growth strategy. Stated differently, it causes the tree to grow by increasing the size of the leaf which reduces loss the most. This method may produce deeper trees with fewer nodes, increasing efficiency and accuracy. Besides this, an additional technique that Light uses to increase speed and efficiency is a histogram-based algorithm that divides continuous feature values into discrete bins. This makes it easier to identify the best-split sites inside the trees, which greatly accelerates training, especially for huge datasets(Aziz et al. 2020).

Decision Tree

A popular machine-learning approach for many classification and regression tasks is the decision tree. A decision tree breaks down a complex problem into smaller alternatives that build up to a conclusion or forecast in an attempt to mimic how people actually make decisions. In the tree structure, a node represents a feature-based choice, a branch represents the decision's outcome, and a leaf represents the final product. The tree begins in step one with a root node that symbolizes the whole dataset (Hatwell et al. 2020). Next, each feature is evaluated by the algorithm to determine which best partitions the data into distinct groups for classification. Based on several selection criteria, including Gini Impurity and Information Gain, the optimal feature to split the node is chosen. In addition, the dataset is subdivided into branches from the root node when the best feature is selected using the previously selected selection criteria. Each branch shows how the feature might turn out. Fresh nodes are created following each branch to reflect decisions made in the context of fresh knowledge. The process of separating data points is carried out recursively until the algorithm either perfectly classifies all data points or meets a predefined stopping threshold. The leaf node that appears after each branch represents the ultimate output, which is a class label in classification (Song et al. 2015).

4. Results and Discussion

The accuracy of two classifiers, LGBM and Decision Tree, is compared in Table I for a range of signal processing and coding strategies. In every test situation, LGBM consistently outperforms the Decision Tree.

Table 4. Results of the Proposed Method

Sr No	Coding Technique	Signal Processing	Classifier	Accuracy (%)
1	Complex	DST	LGBM	94%
2	Complex	Custom Haar Wavelet	LGBM	99%
3	Complex	Haar Wavelet Transform	LGBM	99%
4	Complex	FFT	LGBM	98%
5	EIIP	DST	LGBM	99%
6	EIIP	FFT	LGBM	99%
7	EIIP	Haar Wavelet Transform	LGBM	97%
8	EIIP	Custom Haar Wavelet	LGBM	98%
9	NAN	Haar Wavelet Transform	LGBM	98%
10	NAN	DST	LGBM	98%
11	NAN	FFT	LGBM	98%
12	NAN	Custom Haar Wavelet	LGBM	99%
13	Complex	DST	Decision Tree	92%
14	Complex	Custom Haar Wavelet	Decision Tree	92%
15	Complex	Haar Wavelet Transform	Decision Tree	92%
16	Complex	FFT	Decision Tree	97%
17	EIIP	DST	Decision Tree	99%
18	EIIP	FFT	Decision Tree	98%
19	EIIP	Haar Wavelet Transform	Decision Tree	93%
20	EIIP	Custom Haar Wavelet	Decision Tree	92%
21	NAN	Haar Wavelet Transform	Decision Tree	94%
22	NAN	DST	Decision Tree	97%
23	NAN	FFT	Decision Tree	95%
24	NAN	Custom Haar Wavelet	Decision Tree	95%

Impressively, the accuracies for LGBM often reach 99%, especially when using the Discrete Sine Transform and Custom Haar Wavelet under Complex and EIIP coding schemes. Although useful, the Decision Tree's accuracy is lower, especially when dealing with complex code, where it only reaches 97% accuracy. Notably, EIIP and normalization of atomic number (NAN) coding considerably improve signal characteristics for classification; both

classifiers perform best with these techniques. With EIIP coding, both classifiers perform well, scoring perfect or almost perfect accuracy in cases requiring the Fast Fourier Transform and Discrete Sine Transform.

This pattern indicates that EIIP coding, particularly when combined with particular signal processing methods like wavelets and Fourier transformations, successfully highlights important features for these classifiers. The LGBM classifier is durable and versatile for a variety of data transformations, as evidenced by its outstanding consistency, which keeps it at high accuracy across a wide range of coding and signal processing combinations. While a little less reliable, the Decision Tree's performance is still rather good, especially when it comes to normalizing atomic number (NAN) coding, where it performs exceptionally well when using the Discrete Sine Transform and Haar Wavelet Transform. Although it does so marginally less well than LGBM, this robustness highlights the classifier's ability to handle complicated feature sets. The information as a whole suggests that selecting the appropriate mix of coding and signal processing methods can have a big impact on classification accuracy. While the Decision Tree provides robustness with certain configurations, especially when combined with EIIP and coding, LGBM consistently performs well over a wide range of techniques, making it a better option for complicated datasets. When choosing coding and signal processing techniques for applications that call for complex data analysis and categorization, these insights may be helpful.

4.1 Comparative Analysis of Coding Techniques

Using combined data from LGBM and Decision Tree classifiers, Table 5 offers a consolidated examination of classification accuracies across three coding techniques: complicated, EIIP, and NAN. The complicated technique has the lowest average accuracy, at 95.38%, and the highest standard deviation, at 0.030, suggesting a substantial degree of variability in classifier performance. On the other hand, the EIIP coding method yields better average accuracy of 96.88% and a little smaller standard deviation of 0.026, indicating more consistent outcomes in various configurations. Furthermore, with the lowest standard deviation of 0.017 and an average accuracy of 96.75%, the NAN approach demonstrates comparable performance and yields the most consistent results among the three. While all strategies have the lowest recorded accuracy not going below 92%, NAN normalization has a slightly superior accuracy of 94%. The maximum recorded accuracy is shared by both the EIIP and NAN coding procedures, at 99%. These numbers indicate that although high accuracy may be achieved by all three coding schemes, somewhat better performance and greater reliability are offered by EIIP and NAN when compared to the complicated methodology. This investigation highlights how machine learning models in a variety of signal processing scenarios can benefit from improved predicted accuracy and stability through the use of EIIP and NAN coding approaches.

Table 5. Comparison with coding Technique

Coding Technique	Avg. Acc. (%)	Std. Dev	Worst Acc. (%)	Best Acc. (%)
Complex	95.38	0.030	92	99
EIIP	96.88	0.026	92	99
NAN	96.75	0.017	94	99

4.2 Consolidated Analysis of Signal Processing Techniques

Table 6 offers a thorough examination of how well four signal processing methods perform when applied to various coding strategies and classifiers: DST, Custom Haar Wavelet, Haar Wavelet Transform, and Fast Fourier Transform. For each signal processing method, the average accuracy, standard deviation, worst accuracy, and greatest accuracy are summarized in each row. With a comparatively low standard deviation of 0.027 and the highest average accuracy of 95.50%, the FFT exhibits consistent performance across many contexts. With a standard deviation of 0.033 and an average accuracy of 95.38% and 95.25%, respectively, the Custom Haar Wavelet and Haar Wavelet Transform approaches both indicate a little bit more fluctuation in their performance. Analogously, the DST performs consistently, with an average accuracy of 95.25% and a standard deviation of 0.027, comparable to the FFT. With a range of accuracy from 92% in the worst case to 99% in the best case, all strategies indicate their potential efficacy in certain scenarios. According to the results, all methods work well, but FFT stands out as having somewhat better average accuracy and consistency, which makes it a great option for tasks requiring good performance in signal processing.

Table 6. Comparison with Signal Processing Technique

Signal Processing	Avg. Acc. (%)	Std. Dev	Worst Acc. (%)	Best Acc. (%)
Discrete Sine Transform	95.25	0.027	92	99
Custom Haar Wavelet	95.38	0.033	92	99
Haar Wavelet Transform	95.25	0.033	92	99
Fast Fourier Transform	95.50	0.027	92	99

4.3 Comparative Analysis of Machine Learning Classifiers.

A thorough comparison of the two classifiers' performances—LGBM and Decision Tree—is given in Table 7. With an average accuracy of 98.00%, LGBM performs well, proving its dependability in a variety of situations. With a standard deviation of only 0.014, LGBM is comparatively accurate, indicating that it doesn't change in accuracy according to the coding or signal processing methods employed. LGBM can maintain high accuracy under a variety of scenarios; its peak performance is 99%, while its worst-case accuracy is 94%. Conversely, the Decision Tree classifier displays an average accuracy of 94.67%, which is marginally lower. Its performance appears to be more variable than LGBM, as seen by its greater standard deviation of 0.025. Though its worst-case accuracy falls to 92%, indicating a wider range in performance, the Decision Tree nevertheless attains a 99% best-case accuracy, which is comparable to LGBM. For the datasets and methods under analysis, LGBM is the most reliable classifier since it performs better overall than Decision Tree in terms of average accuracy and consistency. Even if it is a little less reliable, the Decision Tree is still a competitive choice, especially when it attains nearly optimal accuracy.

Table 7. : Comparison among Machine Learning Classifiers

Machine Learning Classifier	Avg. Acc. (%)	Std. Dev	Worst Acc. (%)	Best Acc. (%)
LGBM	98.00	0.014	94	99
Decision Tree	94.67	0.025	92	99

4.4. Comparison with state of art

Table 8 presents a comparative summary of various methodologies applied to the classification of viral diseases, highlighting different approaches and their effectiveness.

Our proposed methodology stands out by classifying subtypes of hepatitis with a high accuracy of 99%, utilizing a combination of coding and advanced signal processing techniques. This task is particularly complex due to the need to differentiate closely related viral subtypes, a challenge not previously addressed with such high accuracy in existing models. In contrast, methods like that of Nilashi et al., while achieving perfect accuracy, are tested on smaller datasets, which may not generalize well to larger, more diverse populations. Additionally, while other approaches relate to our domain, they do not address the specific problem of differentiating between various hepatitis subtypes with such precision. Our approach not only tackles a more nuanced classification challenge but also does so across a substantial dataset, ensuring both robustness and reliability in real-world applications.

Table 8. Comparison with state of art

Sr No	Author Name	Methodology	Dataset	Accuracy (%)
1	Arias et al.	Unsupervised Clustering of DNA Sequence + Deep Learning	Influenza (949), Dengue (1633), Hepatitis B (1562)	99.6
2	Lilhore et al.	Hybrid model for classification using improved SVM and Random Forest	UCI Database consisting clinical data of Hepatitis C	96.67
3	M Fahmy et al.	FCGR + Deep Learning	10 genotypes of Hepatitis C	99
4	Nilashi et al.	Fuzzy Neural Network	155 records of patients suffering from Hepatitis: 32 die (20.6%) and 123 live (79.4%) cases	100
5	Proposed Methodology	3 Coding Techniques + 3 Signal Processing Techniques: Haar Wavelet, Customised Haar Wavelet, FFT	Hepatitis B(1000), Hepatitis C (1000), Hepatitis D (1000), Hepatitis E (1000)	99

This significant advancement sets a new benchmark in the precise application of machine learning to complex medical diagnostic tasks.

5. Conclusion

This study evaluated the effectiveness of signal processing and coding methodologies with LGBM and Decision Tree classifiers. The results demonstrate the dependability of LGBM in many settings, as it continuously outperformed the Decision Tree with an average accuracy of 98.00% and negligible variation. With an average accuracy of 94.67%, the Decision Tree showed more variability despite its usefulness. FFT was shown to be the most efficient signal processing technique, with an average accuracy of 95.50%. Coding strategies like EIIP and NAN normalization, with best-case accuracies of 99%, yielded the best results when paired with FFT and LGBM. The combination of LGBM with FFT and EIIP/NAN normalization coding has to be the most accurate and consistent approach for classification jobs in signal processing.

References

- Tu, T., Patel, K., Shackel, N. A., Viral hepatitis, *Elsevier eBooks*, pp. 317–340, 2017.
- Dougherty, E., Cai, X., Huang, Y., Kim, S., Yamaguchi, R., *Genomic Signal Processing*, Current Genomics, vol. 10, no. 6, p. 364, September 2009.
- Arias, P. M., Alipour, F., Hill, K. A., Kari, L., DeLUCS: Deep learning for unsupervised clustering of DNA sequences, *PLoS ONE*, vol. 17, no. 1, p. e0261531, January 2022.
- Lilhore, U. K., Jagan Mohan, C. H., Bansal, D., Arora, S., Khanna, A., Prashar, D., Arora, S., Suman, S., Hybrid model for precise hepatitis-C classification using improved random forest and SVM method, *Scientific Reports*, vol. 13, no. 1, August 2023.
- Fahmy, A. M., Hammad, M. S., Mabrouk, M. S., Al-Atabany, W. I., On leveraging self-supervised learning for accurate HCV genotyping, *Scientific Reports*, vol. 14, no. 1, July 2024.
- Mueller-Breckenridge, A. J., Dalrymple, J., Cochrane, D., Wilson, K., Molyneux, G., Barnetson, R., MacLean, C., Machine-learning based patient classification using Hepatitis B virus full-length genome quasispecies from Asian and European cohorts, *Scientific Reports*, vol. 9, no. 1, December 2019.

- Mamdouh, H., Shams, M. Y., El-Hafeez, T. A., Hepatitis C Virus Prediction Based on Machine Learning Framework: a Real-world Case Study in Egypt, *Research Square*, January 2022.
- Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., Akbari, E., A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique, *Journal of Infection and Public Health*, vol. 12, no. 1, pp. 13–20, January 2019.
- Sahu, S. S., Panda, G., Identification of Protein-Coding regions in DNA sequences using a Time-Frequency filtering approach, *Genomics, Proteomics and Bioinformatics*, vol. 9, no. 1–2, pp. 45–55, April 2011.
- Akhtar, M., Epps, J., Ambikairajah, E., Signal Processing in Sequence Analysis: Advances in Eukaryotic Gene Prediction, *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 310-321, June 2008.
- Britanak, V., Yip, P. C., Rao, K. R., Fast DCT/DST algorithms, *Elsevier eBooks*, pp. 73–140, 2007.
- Van Fleet, P. J., THE HAAR WAVELET TRANSFORMATION, *Discrete Wavelet Transformations: An Elementary Approach with Applications*, Wiley, pp. 125-181, 2019.
- Hassanzadeh, M., Shahrrava, B., Linear Version of Parseval's Theorem, *IEEE Access*, vol. 10, pp. 27230-27241, 2022.
- Klema, V., Laub, A., The singular value decomposition: Its computation and some applications, *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164-176, April 1980.
- Schmidt, D., A Survey of Singular Value Decomposition Methods for Distributed Tall/Skinny Data, *2020 IEEE/ACM 11th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (Scala)*, GA, USA, pp. 27-34, 2020.
- ML-LGBM, A Machine Learning Model Based on Light Gradient Boosting Machine for the Detection of Version Number Attacks in RPL-Based Networks, *IEEE Journals and Magazine IEEE Xplore*.
- Aziz, N., Akhbar, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., Abas, A. N. C., A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems, *2020 International Conference on Computational Intelligence (ICCI)*, Bandar Seri Iskandar, Malaysia, pp. 11-16, 2020.
- Hatwell, J., Gaber, M. M., Azad, R. M. Atif, Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences, *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, October 2020.
- Song, Y.-Y., Lu, Y., Decision tree methods: applications for classification and prediction, *PubMed Central (PMC)*, April 2015.

Biographies

Love Fadia received his Bachelor of Engineering degree in Electronics and Communication from Vishwakarma Government Engineering College, India in 2022. He is currently an engineering student pursuing his master's degree in Electrical and Computer Engineering at the University of Windsor, Canada. His research interests are: Machine Learning, Deep Learning, Image Processing, Signal Processing and Computer Vision.

Vatsal Shah is an engineering student currently pursuing his master's degree in Electrical and Computer Engineering at the University of Windsor, Canada. He received his Bachelor of Technology in Electronics and Communication from Birla Vishvakarma Mahavidyalaya Engineering College, India in 2021. His research interests are: Machine Learning, Deep Learning, Image Processing and Signal Processing.

MOHAMMAD HASSANZADEH (Member, IEEE) received the B.A.Sc. degree in mathematics from the University of Tehran, in 2000, and the M.A.Sc. and Ph.D. degrees in mathematics from the University of Western Ontario, Canada, in 2006 and 2010, respectively. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Windsor, Canada. His research interests include machine learning, signal processing, time series analysis, image processing, and mathematical methods and transformations in engineering.

Majid Ahmadi received the B.Sc. degree in electrical engineering from Sharif University, Tehran, Iran, in 1971 and the Ph.D. degree in electrical engineering from Imperial College of Science, Technology and Medicine, London, U.K., in 1977. He has been with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, Ontario, Canada, since 1980, currently as Distinguished University Professor. He served Associate Dean of Engineering for Research and Graduate Studies from 2013 to 2023. His research interests include digital signal processing, machine vision, pattern recognition, neural network architectures, applications, and VLSI implementation, computer arithmetic, and MEMS. He has co-authored the book *Digital Filtering in 1-D and 2-Dimensions; Design and*

Applications (New York: Plenum, 1989) and has published over 640 articles in these areas. He is a Fellow IET (U.K) and Fellow IEEE (USA).

Q. M. Jonathan Wu (M'92-SM'09) received a PhD in Computer Vision and Intelligent Systems from the University of Wales, UK. Dr. Wu is a Distinguished professor of electrical and computer engineering and a former Tier 1 Canada Research Chair in Automotive Sensors and Information Systems. He is the founding director of the Computer Vision and Sensing Systems Laboratory at the University of Windsor, Canada. Prior to joining the university, Dr. Wu was a senior research officer at the National Research Council of Canada. He has published one book in the area of 3D computer vision and more than 350 peer-reviewed papers, including 200 journal articles, in the areas of computer vision, machine learning, and intelligent systems. Dr. Wu is/was an associate editor for IEEE Transactions on Cybernetics, IEEE Transactions on Circuits and Systems for Video Technology, and IEEE Transactions on Neural Networks and Learning Systems. He is an elected fellow of the Canadian Academy of Engineering.