2025 10th International Conference on Machine Learning Technologies (ICMLT 2025), Helsinki, Finland, May 23-25, 2025. Final accepted manuscript.

NARX-MLP: A Hybrid Model for Accurate Interpretable Medical Data Classification

Bo Sun
Department of Automatic Control and Systems Engineering
School of Electrical and Electronic Engineering
The University of Sheffield,
Sheffield, UK
bsun14@sheffield.ac.uk

Guoliang Wang
Department of Automatic Control and Systems Engineering
School of Electrical and Electronic Engineering
The University of Sheffield,
Sheffield, UK
gwang42@sheffield.ac.uk

Hua-Liang Wei* [1,2]
[1] *Department of Automatic Control and Systems Engineering, School of Electrical and Electronic Engineering*
[2] *Centre of Machine Intelligence*
The University of Sheffield,
Sheffield, UK
w.hualiang@sheffield.ac.uk

***Abstract*—Machine learning plays a vital role in healthcare, yet medical datasets pose challenges such as nonlinear relationships, high-dimensional features, and the needs for model and result interpretability. We propose an adaptive NARX-MLP classifier, combining NARX with MLP and an adaptive feature selection procedure using L1 regularization. The performance (e.g., accuracy, precision, recall, and F1-score.) of the proposed methods is tested on two datasets: Hepatitis (static) and EEG Eye State (dynamic), to show the superiority of the new method. The selected features by this method can review nonlinear and temporal dependencies and therefore guarantee the capture of complex patterns while maintaining interpretability.**

***Keywords*—Machine learning, NARX, MLP, Feature selection, Classification.**

I. Introduction

Machine learning plays a crucial role in healthcare applications, enabling early diagnosis, disease classification, and patient monitoring. However, medical datasets pose unique challenges, such as nonlinear relationships, high-dimensional feature spaces, and the need for explainability. In particular, traditional machine learning methods often struggle to effectively capture complex dependencies in time-series data like electroencephalography and static datasets such as electronic health records, due to their inherent limitations in modeling nonlinear and high-dimensional relationships between time sequential data[1, 2].

A promising approach to handling nonlinear and time-dependent data is the Nonlinear AutoRegressive with eXogenous inputs (NARX) model [3], which has been widely used in dynamical system modelling problems. NARX effectively captures long-term dependencies and nonlinear interactions by generating polynomial-based features from previous time steps and exogenous inputs. However, despite its success in dynamical regression tasks, its application in solving classification problems remains underexplored, particularly in the healthcare domain. One key challenge is that the expanded feature space introduced by NARX can lead to redundancy and high dimensionality, making it difficult to determine which features contribute meaningfully to classification tasks. Furthermore, feature selection mechanisms traditionally used in NARX-based modelling do not translate well to classification problems, as they are often designed to optimize continuous output rather than discrete decision boundaries.

To address these limitations, we propose a novel approach that integrates NARX with neural networks for improved classification performance. Specifically, we introduce an adaptive NARX-MLP classifier, which combines NARX-based feature generation with a multi-layer perceptron (MLP) to enhance classification accuracy and feature selection. Given the high-dimensional feature space introduced by NARX, we incorporate L1 regularization for sparsity enforcement and introduce the Penalized Error-to-Signal Ratio (PESR) [4] as a stopping criterion to adaptively determine an optimal feature subset. This approach not only balances model complexity and classification accuracy but also reduces the need for extensive manual tuning, making it more practical for real-world healthcare applications. The contributions of this paper are as follows:

- Adaptive Feature Selection—We developed a forward feature selection strategy using L1 regularization and the Penalized Error-to-Signal Ratio (PESR), which adaptively determines the optimal subset of features, reducing redundancy and improving interpretability.
- Adaptive NARX-MLP Classifier—We introduce a novel classifier that extends NARX from regression to classification by integrating it with a multi-layer perceptron (MLP), enabling the effective modeling of nonlinear dependencies in medical datasets while addressing the challenges of high-dimensional feature spaces.
- Improved Stability and Adaptability—Our approach ensures stable and automated feature selection across datasets and is applicable to both static and dynamic data, enhancing generalizability and practicality for real-world healthcare applications.

The remainder of this paper is organized as follows. Section 2 reviews related work on feature selection techniques and classification approaches in medical applications. Section 3 presents the proposed methodology and framework. Section 4

* The corresponding author HLW gratefully acknowledges that this work was supported in part by the Royal Society (Grant Ref: IES-R3-183107).

provides experimental evaluations on both time-series and non-time-series medical datasets, and Section 5 concludes the paper with future research directions.

II. Related Work

Feature selection is usually a critical process in handing high-dimensional medical datasets, with existing methods falling into filter, wrapper, and embedded categories [5]. Filter methods, such as mutual information and correlation-based selection, evaluate features independently of the classifier but may ignore feature interactions. Wrapper methods, including recursive feature elimination and genetic algorithms, optimize feature selection through iterative model evaluations but are computationally expensive. Embedded methods, like L1 regularization, integrate selection within learning but can be unstable in correlated feature spaces.

Machine learning has been widely applied in medical classification, with traditional models such as support vector machines (SVMs), random forests, and XGBoost demonstrating strong predictive performance in static datasets [6, 7]. However, these models struggle with high-dimensional data and often require manual feature selection. Deep learning models, including multi-layer perceptrons (MLPs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), have shown promise in handling complex patterns in medical data [8]. While CNNs excel in image-based classification and RNNs effectively capture sequential dependencies, they typically require large labeled datasets and lack intrinsic feature selection mechanisms. Additionally, the interpretability of these models remains a significant challenge, especially in clinical settings where transparency is critical.

The nonlinear autoregressive model with exogenous inputs (NARX) has been widely used in nonlinear system identification tasks such as time-series forecasting, due to its ability to capture nonlinear dependencies and long-term temporal patterns [9]. Unlike traditional classifiers, NARX can incorporate historical dependencies into feature representations, making it a promising tool for sequential data classification. However, applying NARX to classification introduces new challenges. Its feature expansion can lead to high-dimensional representations, increasing redundancy and complexity. Moreover, NARX lacks an effective feature selection mechanism for classification tasks, which is critical for identifying relevant features and improving interpretability.

To address these challenges, we explore the potential of integrating NARX with neural networks, starting with a lightweight model. Multi-layer perceptrons (MLPs) provide a natural starting point due to their simplicity, efficiency, and ability to learn nonlinear decision boundaries while maintaining interpretability. By combining NARX-generated features with an MLP classifier, we aim to assess whether neural networks can effectively leverage the structured feature representation of NARX while avoiding the computational burden associated with more complex architectures. As the feature space expands, an effective feature selection mechanism becomes essential to prevent overfitting and enhance efficiency. This motivates our proposed adaptive NARX-MLP classifier, which incorporates L1 regularization and the Penalized Error-to-Signal Ratio (PESR) for adaptive and structured feature selection [4] and are presented in the next section.

III. Methodology

A. NARX

NARX is widely used in modeling dynamic systems due to its ability to capture nonlinear dependencies and long-term interactions. NARX has many variants, including neural network-based NARX, autoregressive moving average NARX, and polynomial NARX etc. In this study, we adopt polynomial NARX, which models system dynamics using polynomial transformations of past inputs and outputs.

Polynomial NARX is preferred due to its several attractive features, e.g., explicit feature representation, computational efficiency, and compatibility with structured feature selection. Unlike NN-NARX, which relies on hidden representations within a neural network, polynomial NARX produces interpretable polynomial terms that can be directly analyzed and selected. Additionally, it avoids the high computational cost of training deep architectures during feature generation, making it well-suited for structured classification tasks.

Given an input sequence, polynomial NARX expands the feature space by incorporating past outputs and external inputs through polynomial transformations. The feature vector $\varphi(k)$ at time step k is defined as [10]:

$$\varphi(k) = [y(k-1), \dots, y(k-n_y), u_1(k-d), \dots, u_1(k-d-n_u), \dots, u_r(k-d), \dots, u_r(k-d-n_u)] \quad (1)$$

where y and u_i (i = 1,2…, r) are the system output and the ith input; n_y and n_u are the maximum lags for the ~~system~~ output and inputs; and d is the response time delay between the ~~system~~ output and inputs (usually $d = 0$ or 1).

Expanding these terms using polynomial transformations results in the feature set Φ:

$$\Phi(k) = [\phi_1(\varphi(k)), \phi_2(\varphi(k)), \dots, \phi_M(\varphi(k))] \quad (1)$$

$$M = \binom{n_y + n_u + l}{l} = \frac{(n_y+n_u+l)!}{(n_y+n_u)!\cdot l!} \quad (2)$$

where $\phi_m(\varphi(k))$ represents a polynomial term, l is the nonlinear degree, and M is the total number of polynomial features. This expansion enhances the expressive power of the model but introduces high dimensionality and redundancy, necessitating an effective feature selection mechanism. While originally designed for time-series modeling, polynomial NARX can also be applied to non-sequential data by considering feature interactions instead of temporal dependencies.

B. MLP

The NARX model uses expanded features to capture nonlinear dependencies and potential interactions between variables. However, these features can be high-dimensional and redundant, which may introduce unnecessary complexity into

the classification process. To efficiently leverage these features while ensuring robust classification, we employ a multi-layer perception with L1 regularization as the classifiers.

The expanded feature set Φ serves as the input to the MLP classifier, the MLP input layer is formulated as.

$$z^{(1)} = w^{(1)}\Phi + b^{(1)} \tag{3}$$

where $w^{(1)}$ represents the weight matrix for the input layer and $b^{(1)}$ is the bias vector. The hidden layers apply nonlinear transformations using activation functions such as the Rectified Linear Unit (ReLU), allowing the network to learn hierarchical feature representations [11]. The transformation at the lth hidden layer is defined as:

$$z^{(l)} = f(w^{(l)}z^{(l-1)} + b^{(l)}) \tag{4}$$

where f represents the activation function, $w^{(l)}$ and $b^{(l)}$ denote the weight matrix and bias vector at layer l, respectively. In an MLP network with L layers, the final output layer employs the SoftMax function to compute class probabilities for multi-class classification, given by:

$$\hat{y} = softmax(w^{(L)}z^{(L-1)} + b^{(L)} \tag{5}$$

To enhance feature selection, we apply L1 regularization exclusively on the input layer weights $w^{(1)}$ [12]. This encourages sparsity in the feature space, allowing the MLP model to automatically eliminate less relevant NARX-generated features while retaining those that contribute most to classification. The loss function incorporating L1 regularization is formulated as:

$$L = L_{\text{data}} + \lambda_{\text{L1}} \sum_{\text{i=1}}^{M} \mid W_{\text{i}}^{(1)} \mid \tag{6}$$

where L_{data} is the cross-entropy loss, and λ_{L1} controls the strength of L1 regularization.

C. Forward Selection

After applying L1 regularization, we obtain a reduced set of candidate features by removing those with minimal contribution. However, L1-based selection alone may retain redundant or weakly relevant features. To further refine the selection, we adopt forward feature selection [13], which iteratively evaluates the contribution of each feature using the Penalized Error-to-Signal Ratio (PESR) [4]. PESR incorporates a penalty term to balance model complexity and prediction accuracy, where cross-entropy loss serves as the evaluation metric for prediction error. The process terminates if adding a new feature fails to reduce PESR. The PESR is defined as:

$$PESR_k = \left(\frac{N}{N-\lambda k}\right)^2 \times Loss_k \tag{7}$$

where λ is a parameter controlling the penalty term (which is usually chosen to be $\lambda \geq 1$), and $Loss_k$ is the cross-entropy loss at the kth iteration.

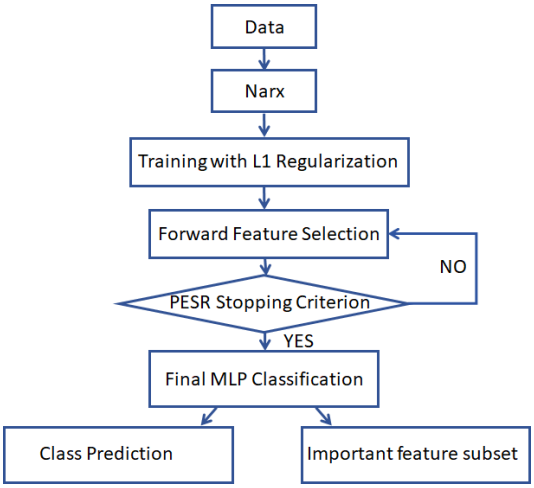

Fig. 1. Framework of NARX-MLP

D. Framework

The proposed Adaptive NARX-MLP classifier integrates NARX-based feature expansion, adaptive feature selection, and MLP-based classification into a unified framework. A key innovation is the combination of NARX-generated polynomial features with structured feature selection, ensuring that only the most relevant features contribute to classification. While NARX effectively captures complex dependencies, its high-dimensional expansion can introduce redundancy. To address this, we apply L1 regularization to the MLP input layer, enforcing sparsity and identifying important features. Further optimization is achieved through forward selection with the PESR index, dynamically balancing model complexity and classification performance. This framework maintains interpretability and adaptability, making it particularly effective for medical classification tasks where feature importance and decision transparency are critical. The overall workflow is illustrated in Figure 1.

IV. Case Studies

A. Datasets

To evaluate the effectiveness of the proposed Adaptive NARX-MLP classifier, we conduct experiments on two medical datasets: the Hepatitis dataset (static) [14] and the EEG Eye State dataset (time-series) [15]. The Hepatitis dataset, sourced from the UCI Machine Learning Repository, consists of 19 clinical features from 615 patients, categorized into five classes: blood donors, suspect blood donors, hepatitis, fibrosis, and cirrhosis. As a static dataset with well-defined clinical features, it serves as an ideal benchmark for evaluating feature selection and classification performance in non-temporal medical data. The EEG Eye State dataset contains EEG recordings from the Emotive EEG Neuroheadset, designed for binary classification of eye state (open or closed). The dataset includes signals from multiple EEG channels and follows a chronological order, making it well-suited for evaluating NARX's ability to capture temporal dependencies.

Our goal is to demonstrate that our approach generalizes effectively across both static and time-series medical

TABLE I. PERFORMANCE COMPARISON

Dataset	Metrics	NARX-MLP	KNN	SVM	RF	CNN	LSTM
Hepatitis	**Accuracy**	**0.994**	0.977	0.969	0.980	0.983	0.948
	Precision	**0.983**	0.946	0.867	0.949	0.898	0.907
	Recall	**0.998**	0.925	0.923	0.958	0.850	0.862
	F1-Score	**0.990**	0.924	0.875	0.952	0.837	0.852
	Specificity	**0.998**	0.979	0.994	0.996	0.982	0.989
EEG Eye State	**Accuracy**	**0.862**	0.770	0.649	0.784	0.741	0.758
	Precision	**0.862**	0.768	0.649	0.786	0.739	0.763
	Recall	**0.859**	0.768	0.628	0.777	0.738	0.750
	F1-Score	**0.856**	0.768	0.625	0.779	0.739	0.752
	Specificity	**0.859**	0.768	0.628	0.777	0.738	0.750

classification tasks. The following sections describe the experimental setup and classification results in detail.

B. Experimental Setup

The datasets were randomly split into 80% training and 20% testing sets. To ensure robust evaluation, we applied 10-fold cross-validation. The NARX model was configured with a nonlinearity degree of 2 to capture higher-order interactions. For EEG Eye State dataset, we incorporated past time steps (lag = 3) to model short-term dependencies in EEG signals.

We evaluated the classification performance using Accuracy, Precision, Recall, F1 Score, and specificity [16]. Accuracy measures overall correctness, while Precision and Recall assess classification reliability. The F1 Score provides a balanced evaluation by combining Precision and Recall, and specificity evaluates the model's ability to correctly identify negative instances.

We compared our Adaptive NARX-MLP Classifier against traditional and deep learning models, including k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), convolutional neural network (CNN) and long short-term memory (LSTM). These baselines help assess the effectiveness of NARX-based feature expansion and PESR-driven feature selection.

C. Results

Table 1 presents the classification performance of the proposed Adaptive NARX-MLP Classifier compared to baseline models on the Hepatitis (static) and EEG Eye State (dynamic) datasets. NARX-MLP consistently outperforms all baseline methods across accuracy, precision, recall, F1-score, and specificity. On the Hepatitis dataset, NARX-MLP achieves the highest accuracy (0.994), significantly outperforming CNN (0.983), LSTM (0.948), and traditional classifiers. The high recall (0.998) and specificity (0.998) highlight the effectiveness of L1-based feature selection and PESR optimization in identifying relevant features without compromising classification confidence. On the EEG Eye State dataset, NARX-MLP achieves an accuracy of 0.862, outperforming CNN (0.741) and LSTM (0.758), demonstrating its ability to effectively model sequential patterns. These results confirm that integrating NARX with MLP and adaptive feature selection significantly improves medical classification performance across both static and dynamic datasets.

To visually compare the performance of the proposed Adaptive NARX-MLP Classifier with baseline models, we present radar charts for the Hepatitis and EEG Eye State datasets (Fig. 2 and Fig. 3). These charts illustrate key metrics accuracy, precision, recall, F1-score, and specificity providing a comprehensive view of each model's performance. The radar charts highlight the superior and balanced performance of NARX-MLP across all metrics, demonstrating its effectiveness in handling both static and dynamic medical datasets. This visualization underscores the robustness of our approach in medical classification tasks.

The adaptive feature selection process identified 11 important features for the Hepatitis dataset and 20 for the EEG Eye State dataset, including nonlinear combinations and time-lagged terms. As table 2 shows, for Hepatitis, key features such

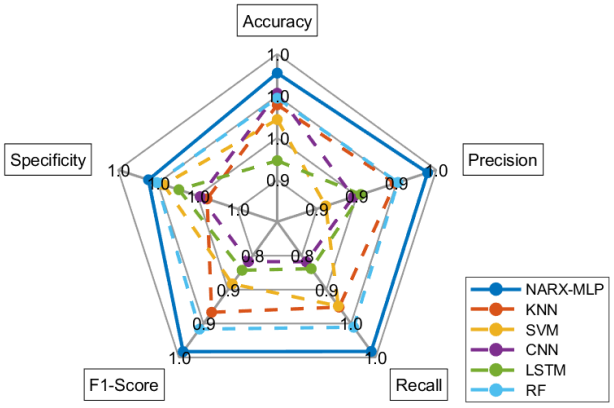

Fig. 2. Radar chart of performance on Hepatitis dataset

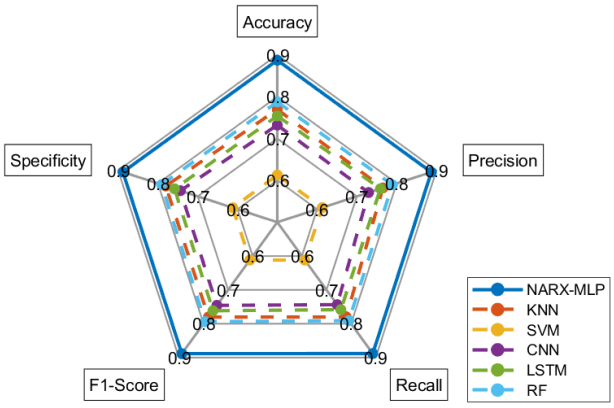

Fig. 3. Radar chart of performance on EEG Eye State dataset.

TABLE II. IMPORTANT FEATURES SELECTED BY NARX-MLP

Description	Hepatitis	EEG Eye State
Important Features (From top to bottom, the importance decreases)	[AST]	[AF3]×[F8(t-1)]
	[CREA]×[GGT]	[F8(t-1)]×[F8(t-3)]
	[ALT]×[CREA]	[F4]×[F4(t-2)]
	[ALP]×[CREA]	[FC6]
	[ALB]×[CHOL]	[O2(t-3)]×[P7(t-3)]
	[ALB]×[ALP]	[O2]×[O2(t-1)]
	[AST]×[CHOL]	…
	[ALP]×[CHE]	[F3(t-2)]×[P8(t-2)]
	[Age]×[BIL]	[AF4(t-1)]×[FC6(t-2)]
	[ALT]×[BIL]	[O2(t-2)]×[P8(t-3)]
	[CHOL]	[F3(t-1)]×[P8(t-1)]

as [AST], [CREA]×[GGT], and [ALB]×[CHOL] were selected, highlighting the significance of biochemical interactions. In EEG Eye State, time-lagged features like [AF3]×[F8(t-1)] and [F4]×[F4(t-2)] were prioritized, demonstrating the model's ability to capture temporal dependencies in dynamic data. These results underscore the effectiveness of our adaptive feature selection mechanism in handling both static and dynamic medical datasets.

D. Discussion

The case studies on the Hepatitis and EEG Eye State datasets demonstrate the effectiveness of the Adaptive NARX-MLP Classifier in handling both static and dynamic medical data. For the Hepatitis dataset, the model successfully identified key biochemical interactions, achieving high accuracy and recall. For the EEG Eye State dataset, it effectively captured temporal dependencies through time-lagged features, outperforming traditional and deep learning baselines. These results highlight the robustness and adaptability of our approach, showcasing its potential for improving medical classification tasks.

A key strength of our framework lies in its integration of NARX-based feature expansion with adaptive feature selection, enabling interpretable and efficient classification. However, the model's performance depends on the quality of input features, and the feature expansion process may introduce computational overhead for high-dimensional datasets. Future work will focus on extending the framework to multi-modal data and optimizing computational efficiency for real-time clinical applications.

V. CONCLUSION

In this study, we proposed the Adaptive NARX-MLP Classifier, a novel approach that integrates NARX-based feature expansion with neural networks to address the challenges of medical classification tasks. Our method demonstrated superior performance on both static and dynamic datasets, outperforming traditional and deep learning baselines across key metrics. The adaptive feature selection mechanism, leveraging L1 regularization and PESR, effectively identified relevant features, including nonlinear combinations and temporal dependencies, enhancing model interpretability and robustness. While this work explores the potential of combining NARX with neural networks, several limitations remain, such as the need for further optimization of hyperparameters and scalability to larger datasets. Future work will focus on refining the model architecture such as attention mechanisms, extending its application to multi-modal datasets, and validating its performance in real-time clinical settings to enhance its practicality and generalizability.

ACKNOWLEDGMENT

The authors gratefully acknowledge that this work was supported in part by the Royal Society (Grant Ref: IES-R3-183107).

REFERENCES

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics,* vol. 19, no. 6, pp. 1236-1246, 2018.

[2] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE journal of biomedical and health informatics,* vol. 22, no. 5, pp. 1589-1604, 2017.

[3] H.-L. Wei, "Sparse, interpretable and transparent predictive model identification for healthcare data analysis," in *Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part I 15*, 2019: Springer, pp. 103-114.

[4] H.-L. Wei and S. A. Billings, "Generalized cellular neural networks (GCNNs) constructed using particle swarm optimization for spatio-temporal evolutionary pattern identification," *International journal of Bifurcation and Chaos,* vol. 18, no. 12, pp. 3611-3624, 2008.

[5] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *bioinformatics,* vol. 23, no. 19, pp. 2507-2517, 2007.

[6] B. Sun and H.-L. Wei, "Machine Learning for Medical and Healthcare Data Analysis and Modelling: Case Studies and Performance Comparisons of Different Methods," in *2022 27th International Conference on Automation and Computing (ICAC)*, 2022: IEEE, pp. 1-6.

[7] M. A. Fayez and S. Kurnaz, "Advanced Hybrid and Preprocessing Models for Diagnosis Challenges in Data Classification," *Journal of Advances in Information Technology,* vol. 15, no. 11, 2024.

[8] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60-88, 2017.

[9] S. A. Billings, "*Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*." John Wiley & Sons, 2013.

[10] H.-L. Wei and S. A. Billings, "Feature subset selection and ranking for data dimensionality reduction," *IEEE transactions on pattern analysis and machine intelligence,* vol. 29, no. 1, pp. 162-166, 2006.

[11] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807-814.

[12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology,* vol. 58, no. 1, pp. 267-288, 1996.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research,* vol. 3, no. Mar, pp. 1157-1182, 2003.

[14] U. M. L. Repository. "HCV Data Set." https://archive.ics.uci.edu/dataset/571/hcv+data (accessed.

[15] J. R. Ayala Solares, H.-L. Wei, and S. A. Billings, "A novel logistic-NARX model as a classifier for dynamic binary classification," *Neural Computing and Applications,* vol. 31, pp. 11-25, 2019.

[16] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management,* vol. 45, no. 4, pp. 427-437, 2009.