

Hankel Imputation for Robust Network Throughput Forecasting

Maria C. Mesquita  [State University of Ceará | clara.mesquita@aluno.uece.br]

Maria L. Linhares  [State University of Ceará | malu.linhares@aluno.uece.br]

Janaina R. Santos  [State University of Ceará | janaina.ribeiro@aluno.uece.br]

Ariel L. Portela  [State University of Ceará | ariel.portela@aluno.uece.br]

Rafael L. Gomes   [State University of Ceará | rafa.lopes@uece.br]

 Center of Science and Technology, State University of Ceará, Fortaleza, CE, 60714-903, Brazil.

Received: 06 November 2025 • **Accepted:** 31 August 2026 • **Published:** 27 October 2026

Abstract Network operators rely on analysis of data of monitoring applications to support strategic planning and ensure Quality of Service (QoS). Among network performance metrics, throughput is particularly critical, as it reflects the actual data successfully delivered between endpoints despite external interference. However, throughput time series are often affected by missing values due to measurement-test failures, the retention of incomplete or failed test records in datasets, and general server-side issues, which compromise data quality and hinder downstream tasks such as traffic forecasting. To address this data quality challenge, we propose HAIL (Hankel Adaptive Imputation via Latent Space), a novel algorithm designed to reconstruct incomplete throughput series using the available throughput measurements and the intrinsic temporal information. HAIL combines Hankel matrix embedding with latent space imputation to capture and preserve the underlying temporal structure. In addition to evaluating HAIL's imputation accuracy, we assessed the downstream utility of the reconstructed datasets by using them to train GRU and LSTM forecasting models and measuring the resulting forecasting performance. The experimental results demonstrate that the HAIL algorithm restores missing throughput patterns, outperforming the baseline methods at missingness rates of 10%, 15%, and 20%, and achieving a slightly lower RMSE in the forecasting results obtained with the datasets imputed by HAIL. Experimental results demonstrate that HAIL effectively restores missing throughput patterns, maintaining temporal consistency and improving the reliability of network performance forecasting.

Keywords: Imputation, Forecasting, Throughput, Hankel, Performance.

1 Introduction

Network throughput is a key metric for assessing the efficiency and capacity of communication systems Younas *et al.* [2023]; Gomes *et al.* [2014]. It quantifies the amount of data successfully transmitted over a network within a specific time interval, reflecting both the system's bandwidth utilization and its ability to sustain service quality under varying traffic conditions Ju and Zhang [2014]; Costa *et al.* [2024]. In network management, throughput is not only a measure of performance but also a critical indicator for resource allocation, congestion control, and Quality of Service (QoS) assurance Arquam and Kumari [2025]; Liu *et al.* [2023]; C. Júnior *et al.* [2022].

However, the reliability of throughput analysis or prediction depends critically on the quality and continuity of the underlying data Gomes *et al.* [2016b]. Throughput time series collected in large-scale research and education networks, such as GÉANT and RNP, are often incomplete because scheduled measurement tests may fail owing to hardware malfunctions, network congestion, or software faults Ferreira *et al.* [2024]; Nobre *et al.* [2026]. Even when tests are completed, the resulting measurements may be inaccurate or inconsistent because of endpoint hardware and operating-system constraints Lopes *et al.* [2023]; Gomes *et al.* [2016a]. Incomplete data often lead to biased or erroneous analyses. Leaving missing values untreated can produce incorrect or ill-defined conclusions in analysis Khayati *et al.* [2020]. Unfortunately, many

data-driven networking studies tend to handle missing data in ad-hoc ways or ignore them altogether Sharma *et al.* [2025]; Pimenta *et al.* [2026], which can undermine the performance of predictive models. Missing or corrupted measurements can distort the interpretation of network behavior, hindering congestion detection, bandwidth estimation, and QoS monitoring Adedayo and Twala [2017]; Nobre *et al.* [2025a].

To address missing values in time series, alternatives such as resampling, discarding incomplete observations, or leaving gaps untreated may be considered; however, these strategies can reduce data availability or compromise subsequent analyses Nascimento *et al.* [2026]. Imputation provides a more suitable alternative by estimating the missing observations from the available data while seeking to preserve the statistical and temporal structure of the original series Khayati *et al.* [2020]; Liu and Li [2022]. This is particularly important for univariate time series, in which trends, seasonality, autocorrelation, and recurrent patterns constitute the main sources of information for downstream tasks such as forecasting Niako *et al.* [2024]; Khattab *et al.* [2023]. The quality of the reconstructed values is commonly assessed by comparing artificially removed observations with their imputed estimates using error-based metrics such as the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) Khayati *et al.* [2020]; Liu and Li [2022]; Niako *et al.* [2024].

Imputation, together with fixed-interval resampling, can improve the consistency and continuity of time-series data, providing a regular temporal representation in which trends,

seasonal behavior, and recurrent patterns can be more effectively analyzed Brito *et al.* [2026]; Gomes *et al.* [2013]. When missing observations are frequent or occur in consecutive intervals, imputation may be necessary to make otherwise incomplete datasets suitable for reliable analysis and modeling Khayati *et al.* [2020]; Khattab *et al.* [2023]; Ferreira *et al.* [2024]. In network-monitoring environments, complete and temporally consistent measurements support network administrators in assessing service performance and maintaining Quality of Service (QoS) and Quality of Experience (QoE) requirements Alkenani and Nassar [2022]; Younas *et al.* [2023]; Adedayo and Twala [2017]. Reconstructed time series may also support downstream analytical tasks, including anomaly detection Zheng *et al.* [2024], time-series clustering Ma *et al.* [2019], and forecasting Niako *et al.* [2024]. In the specific context of network throughput, forecasting is a particularly relevant downstream task because it enables network operators to anticipate performance variations, identify potential bottlenecks, and support proactive resource management Minovski *et al.* [2023]; Portela *et al.* [2024]; Nobre *et al.* [2025b]. Therefore, evaluating forecasting models trained on imputed throughput data provides a complementary means of determining whether an imputation method preserves the temporal information required for subsequent predictive tasks Niako *et al.* [2024]; Wang *et al.* [2024]. Because throughput time series are frequently affected by high missingness rates, measurement failures, and noisy or inconsistent observations Ferreira *et al.* [2024]; Lopes *et al.* [2023], imputation is an important preprocessing step for reconstructing incomplete measurements and enabling otherwise unsuitable datasets to be used in subsequent analysis and forecasting tasks Khayati *et al.* [2020]; Niako *et al.* [2024].

To deal with these challenges, this article introduces a novel approach for imputing missing values in throughput time series, thereby improving the data quality available for analysis and prediction. We propose the Hankel Adaptive Imputation via Latent space (HAIL) algorithm. The core idea is to preserve the intrinsic temporal structures of the throughput series (such as periodic patterns and trends) by projecting the data into a latent space where its fundamental behavior is captured, and then performing intelligent interpolation in that space. In contrast to conventional gap-filling techniques that often oversmooth the data, our methodology leverages an adaptive spectral analysis and low-rank matrix representation (Hankel structure for HAIL) to maintain the shape of the time series.

The imputation process consists of four main stages: (1) Adaptive period detection via Fast Fourier Transform (to estimate dominant periodicity in the series), (2) Matrix embedding of the time series (creating a trajectory Hankel matrix in HAIL), (3) Singular Value Decomposition (SVD) to extract a latent space that captures the series' underlying low-rank structure, and (4) Weighted k -nearest neighbors (KNN) interpolation in the latent space to estimate missing values based on similar temporal patterns. This combination of signal processing and machine learning yields an imputation approach that is tuned to the data's inherent dynamics, aiming to reconstruct missing segments without sacrificing variability or important features.

The benefits of the proposed HAIL algorithm for impu-

tation are demonstrated through experiments that evaluate both reconstruction accuracy and computational resource consumption. Imputation accuracy is quantified using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which are widely adopted for measuring the differences between the original observations and their corresponding imputed estimates Khayati *et al.* [2020]; Liu and Li [2022]; Niako *et al.* [2024]. In addition, the computational feasibility of the evaluated methods is assessed in terms of runtime, CPU time, and peak resident set size (Peak RSS), considering 30 executions for each technique and missingness rate. The experimental results show that HAIL achieves the lowest reconstruction errors among the evaluated methods, with its advantages becoming more evident as the missingness rate increases (Table 2). These findings indicate that HAIL provides a favorable trade-off between reconstruction accuracy and computational cost, producing complete throughput time series that are suitable for subsequent network analysis and forecasting tasks.

Therefore, this article has the following contributions:

- A novel imputation framework for network throughput time series to recover missing throughput measurements while preserving their temporal structures.
- An adaptive period estimation mechanism, enabling the automatic identification of intrinsic periodic behaviors in throughput series and eliminating the need for manually predefined window sizes.
- A latent-space imputation strategy which allows structurally similar temporal patterns to be exploited during the imputation process.
- A comprehensive evaluation using real-world throughput datasets collected from the GÉANT network, demonstrating the applicability of the proposed method in heterogeneous operational scenarios.

The remainder of this article is organized as follows. Section 2 presents the main related studies on network forecasting and data imputation. Section 3 details the proposal, describing the imputation process and the forecasting pipeline. Section 4 describes the datasets, metrics, and experimental setup employed to validate our methodology, as well as discusses the obtained performance results in terms of effectiveness and computational efficiency. Lastly, Section 5 summarizes the main contributions of the work and outlines future research directions.

2 Related Works

The simplest way to handle missing data is to use single imputation strategies that fill in gaps with some plausible values (as opposed to discarding data or using complex multiple-imputation schemes). Traditional single-imputation approaches include methods such as mean or median substitution, forward/backward filling (e.g., last observation carried forward), interpolation (linear, spline, etc.), or moving-average smoothing. These methods are easy to implement and computationally cheap, and they can perform adequately when the missing rate is low or the gaps are very short Khayati *et al.* [2020].

Some smoothing-based methods (exponentially weighted moving averages, Kalman filters, etc.) have also shown robust performance across a range of missing data scenarios Niako *et al.* [2024]. Overall, while basic imputation methods cannot recover fine-grained patterns, they remain strong baselines and are recommended in cases where the time series is univariate, and the gaps are not extensive Niako *et al.* [2024].

Beyond these basics, the literature offers more advanced imputation approaches that seek to exploit the structure in time-series data. Broadly, two classes of methods have been explored: matrix completion techniques and pattern-matching (or proximity-based) techniques Khayati *et al.* [2020]. The matrix completion approach treats the problem as one of reconstructing missing entries in a partially observed matrix, assuming that the true complete matrix has low rank.

In the context of multiple related time series (or a single time series transformed into a matrix form), this means assuming that the temporal dynamics lie near a lower-dimensional subspace. Techniques in this class often construct a large matrix from time-series data and then apply algorithms like SVD, Principal Component Analysis (PCA), or nuclear-norm minimization to fill in the blanks. Low-rank Hankel matrices are a notable example: methods based on Singular Spectrum Analysis (SSA) and its variants fall into this category. For example, Liu *et al.* [2022] propose a Temporal–Spatial Hankel Matrix Factorization to jointly impute long-term geodetic (GNSS) time series from multiple stations Liu and Li [2022]. By exploiting both the temporal correlations (seasonal patterns in each series) and spatial correlations (across nearby sensors), their method was able to recover missing stretches in 10-year coordinate traces with high accuracy Liu and Li [2022]. This illustrates how incorporating the low-rank structure of a time series can greatly improve imputation for datasets with underlying periodic or correlated behavior.

More generally, matrix factorization approaches (including SVD, Nonnegative Matrix Factorization, and robust PCA methods) have been shown to outperform naïve interpolation, especially for block missing patterns, where long contiguous sequences of points are absent Khayati *et al.* [2020]. However, they typically require iterative solving and can be sensitive to parameter tuning (e.g., choosing the rank or regularization terms) Khayati *et al.* [2020].

On the other hand, pattern-matching approaches attempt to fill gaps by finding segments of observed data that resemble the missing portion’s context. These methods do not explicitly assume a global low-rank structure, but rather leverage local similarity in the time series (or across multiple related series). A classic example is *k*-nearest neighbors (KNN) imputation, where for each missing point (or window of points), one finds the most similar historical patterns and uses their values to interpolate the gap.

More sophisticated pattern-based methods use distance measures like Euclidean distance, correlation, or Dynamic Time Warping (DTW) to align and compare time-series subsequences. Dynamic Time Warping is particularly useful for time series because it can measure similarity between sequences that may be phase-shifted or locally stretched/compressed in time. Khattab *et al.* (2023) introduced a pattern-based algorithm called GMA (Gap Matching Algorithm) that exemplifies this approach Khattab *et al.* [2023]. GMA first

determines the dominant period of the time series via Fourier analysis, then for each gap it identifies the most similar “left” and “right” neighboring sequences from the historical data (of length one period) that surround the gap Khattab *et al.* [2023]. By mimicking those neighboring patterns, it reconstructs the missing segment. This method was shown to be effective, particularly for data with strong seasonality. Pattern-matching methods like GMA highlight the importance of preserving local temporal shapes, an aspect sometimes overlooked by global matrix factorization. The drawback, however, is that they can be computationally heavy (DTW is quadratic in sequence length) and may struggle if no close analog of the missing pattern exists in the historical data.

Despite the breadth of research on time-series imputation, there is relatively little work focusing specifically on network throughput or QoS time series. Many network monitoring studies either assume the data is complete or apply off-the-shelf methods without investigating domain-specific improvements. For example, in a recent 5G cellular throughput prediction study, the authors simply applied basic missing-value imputation (using techniques like median filling and one-hot encoding for categorical gaps) as a preprocessing step to clean the dataset Sharma *et al.* [2025]. This helped improve the learning model’s accuracy, but the imputation itself was not the focus and relied on standard tools. In general, network performance data poses some unique challenges: it can exhibit non-stationary behavior, multi-scale seasonalities (daily or weekly usage cycles), and sudden changes due to external events or network reconfigurations. These characteristics suggest that a specialized imputation approach, which can adapt to changing periodicities and preserve bursty events, would be beneficial for network data streams. Our work is motivated by this gap in the literature. We build upon ideas from the aforementioned methods, leveraging the low-rank structure via a Hankel matrix (as in SSA-based methods) and the pattern-matching intuition of neighbor-based interpolation. However, we tailor and combine them in a novel way for the network throughput domain. By adaptively detecting the dominant period of the throughput series and performing imputation in a latent space that captures its principal components, our approach seeks to retain the subtle temporal features (like usage peaks and troughs) that generic methods might smooth out. In the following sections, we detail the proposed HAIL methodology and show through experiments that it addresses the shortcomings of prior approaches, providing more accurate and structure-preserving imputation for throughput time series. Each imputed dataset is then better poised for analysis or prediction tasks, ultimately contributing to more reliable network performance management.

3 Proposal

This section will describe the methodology applied to obtain data for the proposal modeling and the experimental setup, as well as present the HAIL algorithm to perform data imputation under the proposed forecasting pipeline.

3.1 Methodology

This section details the data and the experimental design used for imputation and subsequent forecasting described in Sections 3.1.1 and 3.1.2, respectively.

3.1.1 Throughput data

Throughput was selected as the primary metric in this study due to both its operational relevance and its temporal characteristics. From a network management perspective, throughput is one of the most important QoS indicators, directly supporting tasks such as congestion analysis, SLA verification, resource provisioning, and capacity planning. Moreover, recent advances in predictive network management have increasingly relied on throughput forecasting to enable proactive traffic engineering and autonomous network operation. Consequently, ensuring the quality and continuity of throughput measurements is essential for both operational monitoring and downstream analytical tasks.

From a methodological standpoint, throughput series constitute an appropriate use case for evaluating the proposed HAIL framework because they typically exhibit recurrent daily and weekly patterns, long-term trends, and localized bursty behaviors. These characteristics naturally induce latent low-rank structures and temporal dependencies that can be effectively exploited through Hankel embeddings and latent-space reconstruction techniques. Although this work focuses exclusively on throughput measurements, the proposed methodology is not inherently restricted to this metric and may potentially be extended to other QoS indicators, such as latency, bandwidth availability, and resource utilization. However, metrics characterized by highly irregular or event-driven behaviors, such as packet loss and jitter, may require additional investigation and methodological adaptations, which are left as future research directions.

3.1.2 Methodology and Forecasting Pipeline

The complete experimental workflow adopted in this study is illustrated in Figure 1. The proposed pipeline was designed to evaluate both the effectiveness of the imputation methods and their impact on downstream forecasting tasks under realistic network monitoring conditions.

Initially, raw throughput measurements collected from the GÉANT infrastructure are preprocessed and aggregated into 6-hour intervals to reduce short-term measurement noise and establish a uniform temporal granularity across all datasets. Subsequently, a comprehensive data quality assessment is performed. This stage consists of three complementary steps: (i) periodicity analysis, aimed at identifying temporal cycles within the throughput time series; (ii) missing-data characterization, which quantifies the percentage and distribution of data gaps in each dataset; and (iii) identification of the longest continuous observation interval without missing values, which provides a reliable reference segment for experiments.

Based on this assessment, a two-stage imputation strategy is adopted. First, the original datasets are imputed to reconstruct missing observations and produce complete time series for analysis. Then, to establish a reliable forecasting baseline,

we extract subsets containing only minor data gaps (at most two consecutive missing observations). These subsets are completed using simple interpolation techniques and are considered as reference datasets with minimal distortion. This procedure allows the forecasting experiments to be conducted using high-quality data while simultaneously enabling controlled evaluations under different missing-data conditions.

To investigate the robustness of the forecasting pipeline, artificial missing values are subsequently injected into the reference datasets. Four degraded versions of each dataset are generated, containing 5%, 10%, 15%, and 20% randomly distributed missing observations. These levels were deliberately selected to encompass and slightly exceed the missing rates observed in the original datasets, which ranged from approximately 7% to 18%. Each degraded dataset is then reconstructed using the evaluated imputation methods, including HAIL and the baseline techniques.

The forecasting stage begins with the selection and optimization of the predictive models. Initially, preliminary experiments are performed to identify suitable neural architectures for throughput forecasting. Subsequently, a cross-validation grid search procedure is employed to determine the optimal hyperparameters of the forecasting models, including GRU and LSTM configurations. The imputed datasets generated in the previous stage are then used as inputs for training and evaluating the forecasting models.

Finally, the methodology concludes with a twofold evaluation process. The first component assesses the quality of the imputation methods through reconstruction metrics, quantifying how accurately the original missing values are recovered. The second component evaluates the downstream utility of the reconstructed datasets by measuring the forecasting performance achieved by the neural models trained on the imputed data. This dual evaluation strategy allows us to analyze not only the direct reconstruction capability of HAIL but also its practical impact on predictive network management applications.

3.2 The HAIL Algorithm

Traditional imputation methods often struggle with the non-linear dynamics and complex patterns found in network throughput data, frequently failing to preserve the most important underlying structures. This article addresses this gap by introducing the novel Hankel Adaptive Imputation via Latent Space (HAIL) algorithm, a method designed to impute missing values in network throughput time series. It leverages the inherent low-rank structure of time series data, employing a trajectory matrix (Hankel matrix) to capture complex temporal dynamics. The core strength of the method is the ability to preserve the data's fundamental structure within a latent space and leverage this representation to impute missing values.

The imputation process consists of four sequential stages: first, an adaptive period estimation is performed via Fast Fourier Transform (FFT); second, the series is transformed into a representation matrix, using a Hankel structure (HAIL); third, Singular Value Decomposition (SVD) is used to extract the latent structure of this matrix; and finally, missing values are imputed through a weighted neighbor-based reconstruction within this low-dimensional space. Each stage is

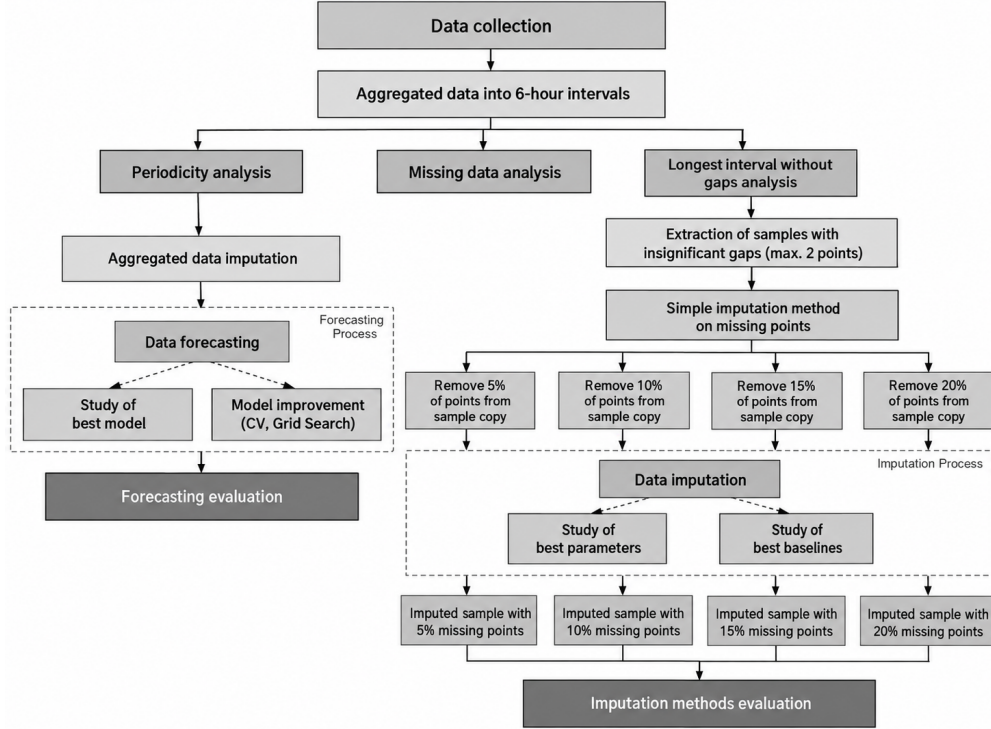


Figure 1. Detailed workflow.

designed to preserve the series' underlying dynamics while mitigating the smoothing artifacts common in traditional imputation methods.

3.2.1 Formulation

This section provides the mathematical formalization of the HAIL imputation method, detailing the intuition and computations required, where Table 1 summarizes the notation applied. Extracting a meaningful latent space using decomposition methods like Singular Value Decomposition (SVD) is not directly possible on a 1D univariate time series. This necessitates a prerequisite step: transforming the 1D series, y , into a 2D representation matrix, M , which is assumed to have a low-rank structure.

The critical first step in this transformation is determining the matrix's optimal dimension, P . This parameter must align with the dominant period of the time series, enabling the matrix to effectively capture the data's underlying cyclical patterns. To identify this period, we employ the Fast Fourier Transform (FFT), a standard and robust method for spectral analysis. Given the pre-processed series y_{norm} of length N , the DFT yields a set of complex coefficients Y_k : $Y_k = \sum_{n=0}^{N-1} y_{norm}(n) \cdot e^{-i \frac{2\pi}{N} kn}$ for $k = 0, \dots, N-1$. From these coefficients, we compute the signal's power spectrum, $S(f)$, as the squared magnitude of the coefficients: $S(f) = |Y(f)|^2$. Frequencies f are then converted to periods $T = 1/f$. The algorithm identifies the optimal period P as the one corresponding to the maximum power in the spectrum, subject to predefined constraints (P_{min}, P_{max}): $P = \arg \max_{T \in [P_{min}, P_{max}]} (S(1/T))$. The resulting value P is the foundational parameter used to construct the representation matrix.

Table 1. Notation Table

Symbol	Description
y	throughput one-dimensional data
M	two-dimensional matrix representation of the throughput time series
P	matrix of optimal dimension
$Nk(i, j)$	length of preprocessed y_{norm}
Y_k	set of complex coefficients
$S(f)$	signal spectrum from coefficients
$e^{-i \frac{2\pi}{N} kn}$	Fourier kernel
H	Hankel matrix
U	left orthogonal matrix
V	right orthogonal matrix
Σ	diagonal matrix of non-negative singular values
si	non-negative singular values
r	rank
$Energy$	cumulative energy threshold
d	Euclidean distance
Nk	neighborhood set
w	weight

Table 2. Imputation Results.

Technique	5%		10%		15%		20%	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
HAIL	6.73	1.08	13.44	3.23	16.55	4.91	20.12	6.79
	± 4.62	± 0.74	± 8.29	± 2.00	± 8.47	± 2.29	± 10.20	± 3.20
HoltWinters	5.47	0.93	16.62	3.86	19.71	5.68	23.83	8.23
	± 4.86	± 0.84	± 10.43	± 2.47	± 10.64	± 2.98	± 12.47	± 4.32
ARIMA	6.26	1.02	17.84	4.26	22.36	6.55	25.74	9.15
	± 4.52	± 0.77	± 11.96	± 3.02	± 11.54	± 3.35	± 12.63	± 4.70
KNN	7.59	1.20	14.29	3.42	17.74	5.23	21.35	7.34
	± 5.06	± 0.76	± 8.02	± 2.01	± 7.90	± 2.20	± 9.81	± 3.37
Kalman	8.17	1.28	163.00	30.17	173.79	38.17	169.94	38.87
	± 4.77	± 0.81	± 159.55	± 37.56	± 164.95	± 41.62	± 136.57	± 30.61
LOCF	9.36	1.45	19.61	4.57	26.64	7.50	30.07	10.20
	± 11.83	± 1.77	± 14.73	± 3.55	± 13.96	± 3.96	± 16.01	± 5.74
Linear	7.98	1.28	16.83	3.96	21.45	6.23	23.84	8.11
	± 5.35	± 0.88	± 10.60	± 2.56	± 10.22	± 2.78	± 11.08	± 3.69
Spline	10.51	1.64	21.74	5.14	31.91	8.77	36.87	12.15
	± 9.08	± 1.42	± 13.95	± 3.26	± 17.17	± 4.60	± 20.42	± 6.72

Since a decomposition method like SVD cannot be applied directly to a 1D time series. We need to transform the one-dimensional network throughput series y into a 2D representation matrix M , which is assumed to have an underlying low-rank structure. Exposing this structure is the key to identifying the latent space, and the idea is to construct a representation matrix H , which has a Hankel structure. The fundamental idea is that the temporal dynamics of a 1D time series conceal a low-rank structure, which becomes explicit when the series is embedded in this higher-dimensional space. This allows SVD to effectively separate the principal dynamic components from noise.

Given the input series $y = (y_1, y_2, \dots, y_n)$ and the window size $L = P$, the representation matrix H is constructed by sliding the window across the series. The resulting matrix has $K = n - L + 1$ rows and L columns, where each row $H_i = (y_i, \dots, y_{i+L-1})$ represents a state vector of L consecutive observations.

Here, our method departs from the traditional energy calculation, which uses the sum of squared singular values (proportional to variance). Based on empirical experiments, we define energy as the direct sum of the singular values. The cumulative energy for a given rank k is calculated as: $\text{cum_energy}(k) = \frac{\sum_{i=1}^k s_i}{\sum_{i=1}^{\min(T,P)} s_i}$, where the rank r is then selected as the smallest k that satisfies the energy threshold as $r = \arg \min_k \{\text{cum_energy}(k) \geq E_{\text{energy}}\}$. This decision to use the direct sum s_i instead of s_i^2 is a critical and deliberate design choice. We found that this approach avoids overemphasizing/overweighting the first few dominant components, which can otherwise mask more subtle patterns. This leads to a more balanced rank selection and ultimately produces more stable projections of the data into the latent space.

With the optimal rank r determined, we construct the latent space representation, Z . This projection is formed using the first r left singular vectors (the rows of U), as these represent the principal components associated with the matrix rows (our

state vectors or periodic cycles). These vectors are weighted by their corresponding singular values s_i , which act as a measure of their importance: $Z = U_{:,1:r} \cdot \Sigma_{1:r,1:r}$. Each row Z_i of this new matrix Z serves as a low-dimensional "summary vector" (or embedding) of the corresponding row M_i from the representation matrix, capturing its essential structural patterns.

The core imputation is then performed using a K-Nearest Neighbors (KNN) approach *within* this low-dimensional latent space. This is a critical step: by calculating distances in Z rather than in the original, high-dimensional matrix M , we find neighbors based on structural similarity (i.e., rows with similar temporal patterns or seasonal shapes) rather than on noisy, point-wise similarities.

The algorithm iterates through the *original* matrix M to find each missing value $M(i, j)$. For each missing entry, the process begins by retrieving the summary vector Z_i for the target row i . Next, it identifies the set of candidate rows $\mathcal{C}$ where the value in the same column j is *observed* ($M(c, j) \neq \text{NaN}$). The Euclidean distance $d(Z_i, Z_c)$ is then computed in the latent space between the target vector Z_i and all candidate vectors Z_c where $c \in \mathcal{C}: d(Z_i, Z_c) = \|Z_i - Z_c\|_2 + \epsilon$ (where ϵ is a small constant, e.g., 10^{-8} , to prevent division by zero). The k candidates with the smallest distances are selected to form the neighbor set $\mathcal{N}_k \subseteq \mathcal{C}$. Finally, the missing value $M(i, j)$ is imputed using an Inverse Distance Weighted (IDW) average of the values from this neighbor set, which assigns greater influence to neighbors that are structurally closer in the latent space: $M_{\text{imputed}}(i, j) = \frac{\sum_{c \in \mathcal{N}_k} w_c \cdot M(c, j)}{\sum_{c \in \mathcal{N}_k} w_c}$, where $w_c = \frac{1}{d(Z_i, Z_c)}$. This process is repeated for every missing value, resulting in a fully imputed representation matrix, M_{imputed} .

4 Results

4.1 GEANT Data Collection

The datasets analyzed in this work originate from the **GEANT Performance Measurement Platform (PMP)**, a pan-European initiative designed to monitor end-to-end network performance across National Research and Education Networks (NRENs). Each dataset corresponds to throughput measurements collected by **perfSONAR** nodes, which are specialized servers distributed across strategic network points that continuously test available bandwidth between domains. **perfSONAR** (performance Service-Oriented Network Architecture) is an open, federated monitoring framework collaboratively maintained by GEANT, ESnet, Internet2, and several NRENs worldwide. Its federated design ensures interoperability and methodological consistency, enabling cross-domain analyses based on homogeneous data formats and measurement protocols such as **iperf3** for throughput and **owamp** for latency and packet loss.

Within the GEANT ecosystem, the **PMP nodes** serve as high-performance reference points positioned in the network core. The datasets employed in this research include measurements from nodes such as **psmp-gn-bw-lis-pt.geant.org** (Lisbon, Portugal), **psmp-gn-bw-poz-pl.geant.org** (Poznan, Poland), and **psmp-gn-bw-vie-at.geant.org** (Vienna, Austria). These nodes operate on dedicated 10–100 Gbps backbone links and follow a standardized naming convention that encodes their function (e.g., **bw** for bandwidth testing), geographic location, and organizational domain. This uniform naming strategy enables automated identification of node characteristics, facilitating topological and geospatial analyses of network performance paths.

Complementary datasets were collected from **National and Regional Research and Education Networks (NRENs and RENs)** that connect universities and scientific institutions to GEANT’s backbone. These include the Turkish node **perfsonar-ankara.ulakbim.gov.tr** operated by TÜBİTAK-ULAKBİM (ULAKNET), the Spanish node **perfsonar-sonda.rediris.es** maintained by RedIRIS, the Luxembourgish node **perfsonar.restena.lu** managed by Fondation Restena, and the Catalan regional node **psmp-anella.csuc.cat** operated by CSUC under the “Anella Científica.” Each of these nodes serves as a national or regional monitoring point, ensuring visibility into local network performance and its interconnection to the continental infrastructure. Finally, the dataset includes a **campus-level node**, **psmall.lut.ac.uk**, deployed at Loughborough University (UK), representing the network’s edge where users and scientific instruments operate. This small-form-factor **perfSONAR** node exemplifies the bottom tier of GEANT’s hierarchical architecture, focusing on intra-campus and metropolitan diagnostics rather than long-haul backbone testing.

Together, these datasets provide a multilevel view of the European research networking ecosystem, spanning institutional, national, and transnational layers. The hierarchical organization of the nodes enables the analysis of throughput patterns across diverse operational contexts, ensuring that the study’s results capture both backbone-level behavior and

local variability inherent to distributed infrastructures.

This section presents a comprehensive evaluation of the proposed HAIL algorithm against several established imputation methods. The analysis is structured in three parts:

1. A direct comparison of imputation accuracy using standard error metrics (RMSE, MAE) across varying data sparsity levels.
2. An assessment of downstream forecasting quality, where a Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) models are trained on the imputed data.
3. A computational performance experiment to analyze memory consumption and execution time.

4.2 Imputation Effectiveness

The experimental results, summarized in Table 2, clearly establish the superiority of the proposed HAIL method over existing techniques. Across all tested percentages of missing data, HAIL consistently achieved the lowest RMSE and MAE. This advantage was not only present in low-sparsity scenarios (e.g., 5% missing data where HAIL achieved a 6% improvement) but became significantly more pronounced as the challenge increased.

Furthermore, HAIL’s design demonstrates superior scalability (Table 2). While traditional approaches like the Kalman filter experienced performance degradation when missing data exceeded 10%, registering unusable error rates, HAIL’s performance degradation was controlled and graceful. Similarly, HAIL consistently surpassed ARIMA in all conditions and outperformed HoltWinters in all scenarios except for the 5% missing data mark (Table 2).

4.3 Forecasting Capacity

The ultimate evaluation of an imputation method is its downstream quality, i.e., the ability of the imputed data to successfully train a forecasting model. As shown in Table 3, when applied to a Gated Recurrent Unit (GRU) forecasting task, our HAIL algorithm delivered the strongest performance, achieving the lowest overall Root Mean Squared Error (RMSE: 66.52). This result surpassed complex statistical methods like ARIMA (69.67) and Holt-Winters (69.96), while methods like Kalman (105.78 RMSE) and Spline (97.23 RMSE) failed catastrophically (Table 3).

Two notable observations emerged from Table 3. First, the failure of sophisticated methods like Kalman suggests their smooth imputations may remove essential data variance. Second, simple methods like Mean (66.82 RMSE) performed exceptionally well, nearly matching HAIL. This indicates that preserving the data’s general distribution (Mean) or local similarity (such as KNN) is highly effective for this task. However, HAIL’s results in RMSE, the metric most sensitive to large errors, is the key point (Table 3). Its superior RMSE demonstrates that its imputations lead to fewer severe forecasting errors, supporting its suitability for a pipeline.

Table 3. GRU and LSTM Forecasting Results.

Imputation Technique	GRU		LSTM	
	RMSE	MAE	RMSE	MAE
HAIL	66.52 ± 23.82	37.53 ± 16.59	66.85 ± 24.91	37.68 ± 16.90
Mean	66.78 ± 25.20	36.94 ± 17.89	67.88 ± 26.42	38.42 ± 19.92
SVD_KNN	68.33 ± 24.77	40.94 ± 19.12	68.34 ± 24.81	40.74 ± 18.54
ARIMA	69.67 ± 28.05	41.20 ± 20.92	71.83 ± 31.12	42.95 ± 23.27
HoltWinters	71.06 ± 28.78	42.68 ± 20.44	73.04 ± 32.06	44.71 ± 23.16
Linear	71.89 ± 27.99	45.26 ± 22.65	75.19 ± 31.97	47.75 ± 25.03
LOCF	77.67 ± 38.27	48.72 ± 26.32	80.04 ± 41.77	50.06 ± 28.38
Spline	97.23 ± 36.33	68.10 ± 32.42	108.40 ± 43.05	75.28 ± 35.78
Kalman	105.78 ± 89.10	68.55 ± 54.27	111.65 ± 104.73	72.31 ± 62.85

4.4 Computational Resources Usage

HAIL’s performance was analyzed across 30 runs for each technique and each missingness rate, where the results show the mean and 95% confidence interval. Table 4 reports three measures of computational resource usage. Runtime represents the time required to complete one imputation process, expressed in seconds per run and averaged across the executions performed for each technique and missingness rate. CPU time measures the time during which the processor was actively executing the imputation procedure, excluding most waiting periods, scheduling delays, and other non-computational overhead. Comparing CPU time with runtime helps determine whether execution time is primarily associated with computation or with external delays. Finally, Peak RSS reports the maximum amount of memory usage by the process during execution, expressed in MiB. This metric is used to assess whether the memory requirements of each technique remain compatible with execution on resource-constrained environments, including personal computers.

Table 4. Computational Performance Comparison.

Technique	Runtime (s/run)	CPU Time (s/run)	Peak RSS (MiB)
HAIL	0.926 ± 0.425	0.897 ± 0.426	164.78 ± 46.41
HoltWinters	6.615 ± 3.196	6.375 ± 3.017	164.78 ± 46.40
ARIMA	3.598 ± 0.997	3.305 ± 0.861	164.78 ± 46.40
KNN	0.248 ± 0.071	0.245 ± 0.080	164.78 ± 46.40
Kalman	1.242 ± 0.384	1.222 ± 0.379	169.99 ± 44.67
LOCF	0.027 ± 0.006	0.017 ± 0.019	164.79 ± 46.40
Linear	0.047 ± 0.012	0.039 ± 0.021	164.78 ± 46.40
Spline	0.226 ± 0.050	0.206 ± 0.051	164.79 ± 46.40
Mean	0.032 ± 0.008	0.025 ± 0.015	164.78 ± 46.40

As shown in Table 4, the HAIL imputation technique is 7.1, 3.9, and 1.3 times faster than more computationally demanding benchmark methods such as Holt–Winters, ARIMA, and Kalman, respectively. Moreover, the comparison with the Kalman-based method is particularly relevant because both methods explicitly model temporal structure through more elaborate numerical procedures than direct replacement or interpolation techniques. However, some simpler methods

outperform HAIL in runtime, which is relatively expected, because they employ lower-complexity procedures, unlike HAIL, which performs adaptive period estimation, Hankel matrix construction, SVD-based latent-space extraction, and repeated neighbor searches for missing-value reconstruction. Although they achieve lower runtimes, these approaches do not model temporal dynamics as extensively as HAIL. Although the two metrics are similar, CPU time provides complementary information because it represents the time spent actively executing computations, excluding most non-computational overhead.

Regarding the peak process memory, it is possible to observe that among the methods under the adopted measurement procedure, there is no statistically significant difference in performance. Besides that, the HAIL method remains with the same average memory usage as the other techniques, including the less complex, highlighting the benefits of the proposal.

These results demonstrate a trade-off: HAIL does not have the best performance of runtime, but it achieves the lowest reconstruction errors. These findings position HAIL as a suitable solution between inexpensive statistical baselines and slower model-based methods for network throughput imputation.

4.5 Final Discussion

Based on the experimental results, the HAIL algorithm can be validated as a superior imputation solution from three distinct perspectives. First, in terms of direct imputation accuracy (Table 2), HAIL outperformed existing methods, with its superiority becoming more pronounced as data sparsity increased. Second, and most critically, HAIL demonstrated the best downstream utility, providing a data structure that enabled a GRU and LSTM models to achieve the lowest forecasting error (Table 3). This confirms its ability to preserve complex temporal dynamics where other methods fall short. Finally, regarding computational cost, HAIL is the most balanced method, since it has the most efficient execution time and resource consumption.

These results indicate the benefits of HAIL in several real-world network operation scenarios where continuous monitoring data are frequently affected by measurement failures, communication interruptions, or telemetry inconsistencies. One particularly relevant application is in Internet Service Providers (ISPs), where network operators continuously

collect throughput measurements from geographically distributed probes and missing observations may compromise long-term capacity planning, SLA verification, and congestion diagnosis. By reconstructing missing measurements while preserving intrinsic temporal patterns, HAIL enables operators to maintain consistent historical datasets and perform more reliable trend analyses, even in the presence of substantial data loss. Additionally, HAIL enables suitable network management and proactive resource allocation, since forecasting models are highly sensitive to data quality and may suffer significant degradation when trained on incomplete or poorly imputed datasets. Because HAIL preserves both global periodic trends and local throughput fluctuations, its adoption can substantially improve the quality of analytics pipelines. Another promising application domain is network anomaly detection and cybersecurity monitoring. Many anomaly detection techniques rely on historical throughput baselines to identify abnormal traffic patterns associated with distributed denial-of-service attacks, equipment failures, routing instabilities, or misconfigurations. Missing observations may distort these baselines, leading to false alarms or missed detections. Because HAIL preserves temporal dynamics and avoids excessive smoothing, it can reconstruct monitoring datasets while retaining anomalous behaviors and traffic bursts essential for security analytics. Overall, these scenarios demonstrate that HAIL is not merely an imputation technique for offline data preprocessing but rather a practical enabler for next-generation network intelligence systems, supporting reliable monitoring, forecasting, anomaly detection, and autonomous network operation in increasingly complex and data-driven communication infrastructures.

5 Conclusion

This article introduced HAIL, a novel approach for reconstructing missing values in network throughput series through a combination of spectral analysis, matrix embedding, and latent space interpolation. By exploiting the low-rank structure of time-series data, HAIL effectively preserved temporal patterns even under high sparsity conditions. Across all evaluated metrics and missing data levels, it consistently outperformed traditional baselines such as ARIMA, Holt-Winters, and Kalman filters, demonstrating robustness and scalability. The downstream forecasting results using GRU and LSTM networks confirmed that HAIL not only improved raw imputation accuracy but also enhanced predictive performance.

Beyond its superior quantitative results, HAIL's design provides valuable insight into the balance between structural preservation and computational efficiency in time-series imputation. The integration of Hankel embedding with latent-space neighbor reconstruction enables the method to preserve both global periodic behaviors and local temporal fluctuations, which is particularly important for datasets characterized by bursty, seasonal, or non-stationary dynamics, such as network throughput measurements. The results demonstrate that HAIL successfully preserves the predictive information contained in the original series while simultaneously achieving superior reconstruction accuracy. This indicates that the method avoids the excessive smoothing effects commonly introduced

by traditional imputation approaches, maintaining temporal characteristics that are essential for subsequent analytical tasks. Furthermore, the practical benefits of HAIL extend beyond forecasting performance alone. The method consistently exhibits robust behavior under increasing levels of data sparsity, providing graceful performance degradation even in scenarios where competing techniques experience substantial deterioration. Such a feature is particularly valuable in real-world network monitoring environments, where measurement failures and telemetry gaps are often unpredictable and heterogeneous. By generating high-fidelity reconstructed datasets, HAIL can support a wide range of network management activities, including anomaly detection, SLA verification, traffic engineering, predictive maintenance, and long-term capacity planning. These findings underscore HAIL's potential as a general framework for restoring incomplete network monitoring data and enabling more reliable network intelligence systems. Rather than providing improvements solely in a single predictive metric, HAIL offers a comprehensive compromise between reconstruction accuracy, temporal consistency, robustness, and downstream utility. Consequently, its adoption can contribute to more dependable forecasting pipelines, more accurate historical repositories of network telemetry, and more informed operational decision-making across complex communication infrastructures.

As a promising avenue for future research, we intend to extend HAIL toward multivariate time-series imputation and forecasting, where inter-metric dependencies (such as correlations among throughput, latency, jitter, and packet loss) can be jointly modeled. Incorporating cross-variable structures through tensor-based Hankel embeddings or graph-regularized latent spaces may further enhance the algorithm's ability to capture both temporal and spatial relationships. Such an extension would pave the way for more comprehensive network intelligence systems capable of simultaneously analyzing multiple QoS dimensions and supporting predictive maintenance and anomaly detection in real-time network operations.

Acknowledgements

The authors would like to thank the CNPq (*N*° 305946/2025-0 and *N*° 405940/2022-0) and CAPES (*N*° 88887.954253/2024-00 and *N*° 88887.972043/2024-00) of Brazil for the financial support.

Funding

The authors would like to thank the CNPq (*N*° 305946/2025-0 and *N*° 405940/2022-0) and CAPES (*N*° 88887.954253/2024-00 and *N*° 88887.972043/2024-00) of Brazil for the financial support.

Authors' Contributions

All authors contributed to the development of the proposal and the writing of this manuscript.

Competing interests

The authors declare no conflicts of interest.

Availability of data and materials

The code developed, as well as the data used in the experiments, are available in the project repository (<https://github.com/LarcesUece/>) with the necessary instructions for reproducibility.

References

- Adedayo, A. O. and Twala, B. (2017). Qos functionality in software defined network. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 693–699. DOI: 10.1109/ICTC.2017.8191068.
- Alkenani, J. and Nassar, K. (2022). Network monitoring measurements for quality of service: A review. *Iraqi Journal for Electrical and Electronic Engineering*, 18:33–42. DOI: 10.37917/ijeee.18.2.5.
- Arquam, M. and Kumari, S. (2025). *Maximization of Communication Network Throughput Using Dynamic Traffic Allocation Scheme*, page 116–127. Springer Nature Singapore. DOI: 10.1007/978-981-96-5887-9_9.
- Brito, M. L. L., Ferreira, M. C. M., Portela, A. L. C., and Gomes, R. L. (2026). Ai-based estimation of bandwidth availability for data offloading in edge-cloud computing. *IEEE Networking Letters*, 8:69–73. DOI: 10.1109/LNET.2025.3614770.
- C. Júnior, E., L. Costa, W., L. C. Portela, A., S. Rocha, L., L. Gomes, R., and M. C. Andrade, R. (2022). Detecting attacks and locating malicious devices using unmanned air vehicles and machine learning. *Journal of Internet Services and Applications*, 13(1):11–20. DOI: 10.5753/jisa.2022.2327.
- Costa, M. A., Costa, Y. M., Almeida, Y. O., Cardoso, F. J., and Gomes, R. L. (2024). Connection management using automated firewall based on threat intelligence. In *Proceedings of the 2024 Latin America Networking Conference, LANC '24*, page 32–37, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3685323.3685331.
- Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7. DOI: 10.23919/CNSM62983.2024.10814394.
- Gomes, R. L., Bittencourt, L. F., Madeira, E. R., Cerqueira, E., and Gerla, M. (2016a). A combined energy-bandwidth approach to allocate resilient virtual software defined networks. *Journal of Network and Computer Applications*, 69:98–106. DOI: 10.1016/j.jnca.2016.02.024.
- Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2013). A virtual network allocation algorithm for reliability negotiation. In *2013 22nd International Conference on Computer Communication and Networks (ICCCN)*, pages 1–7. DOI: 10.1109/ICCCN.2013.6614097.
- Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2014). Supporting sla negotiation for vsdn based on similarity and price issues. In *2014 IEEE 13th International Symposium on Network Computing and Applications*, pages 287–290. DOI: 10.1109/NCA.2014.49.
- Gomes, R. L., Bittencourt, L. F., Madeira, E. R. M., Cerqueira, E. C., and Gerla, M. (2016b). Software-defined management of edge as a service networks. *IEEE Transactions on Network and Service Management*, 13(2):226–239. DOI: 10.1109/TNSM.2016.2538821.
- Ju, H. and Zhang, R. (2014). Throughput maximization in wireless powered communication networks. *IEEE Transactions on Wireless Communications*, 13(1):418–428. DOI: 10.1109/TWC.2013.112513.130760.
- Khattab, A. A. R., Elshennawy, N. M., and Fahmy, M. (2023). Gma: Gap imputing algorithm for time series missing values. *Journal of Electrical Systems and Information Technology*, 10(1):41. DOI: 10.1186/s43067-023-00094-1.
- Khayati, M., Lerner, A., Tymchenko, Z., and Cudré-Mauroux, P. (2020). Mind the gap: an experimental evaluation of imputation of missing values techniques in time series. *Proc. VLDB Endow.*, 13(5):768–782. DOI: 10.14778/3377369.3377383.
- Liu, H. and Li, L. (2022). Missing data imputation in gnss monitoring time series using temporal and spatial hankel matrix factorization. *Remote Sensing*, 14(6). DOI: 10.3390/rs14061500.
- Liu, X., Xu, B., Zheng, K., and Zheng, H. (2023). Throughput maximization of wireless-powered communication network with mobile access points. *IEEE Transactions on Wireless Communications*, 22(7):4401–4415. DOI: 10.1109/TWC.2022.3225085.
- Lopes, R., Rand, D., Chown, T., Golub, I., and Vuletić, P. (2023). Network performance tests over the 100G BELLA link between GÉANT and RNP. White Paper GN4-3-22-T5D0T7, GÉANT Association.
- Ma, Q., Zheng, J., Li, S., and Cottrell, G. W. (2019). Learning representations for time series clustering. Available at: <https://dl.acm.org/doi/abs/10.5555/3454287.3454626#sec-terms>.
- Minovski, D., Ögren, N., Mitra, K., and Åhlund, C. (2023). Throughput prediction using machine learning in lte and 5g networks. *IEEE Transactions on Mobile Computing*, 22(3):1825–1840. DOI: 10.1109/TMC.2021.3099397.
- Nascimento, E. S., Pimenta, I. A., Lee, M. H., de Araujo, T., and Gomes, R. L. (2026). A decoupled embedding-based framework for efficient node classification on social networks. *Journal of Internet Services and Applications*, 17(1):214–223. DOI: 10.5753/jisa.2026.7105.
- Niako, N., Melgarejo, J. D., Maestre, G. E., and Vatcheva, K. P. (2024). Effects of missing data imputation methods on univariate blood pressure time series data analysis and forecasting with arima and lstm. *BMC Medical Research Methodology*, 24(1):320. DOI: 10.1186/s12874-024-02448-3.
- Nobre, F. V. J., Alves, D. O., Araujo, R. S., Campos, G. A., and Gomes, R. L. (2026). Risk classification of ip addresses using machine learning with weighted voting approach. In Rodrigues, L. A. and Oliveira, R., editors, *Dependable and Secure Computing*, pages 320–328, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-032-11539-3_9.
- Nobre, F. V. J., Araujo, R. S., Alves, D. O., Nascimento,

- E. S., Rodrigues, L. S., and Gomes, R. L. (2025a). Network connection management based on the integration of artificial intelligence for it operations and threat intelligence databases. In *2025 23rd International Symposium on Network Computing and Applications (NCA)*, pages 95–102. DOI: 10.1109/NCA67271.2025.00027.
- Nobre, F. V. J., Silva, D. d. S., Ferreira, M. C. M. M., Brito, M. L. M. L., de Araújo, T. P., and Gomes, R. L. (2025b). Time-weighted correlation approach to identify high delay links in internet service providers. *Journal of Internet Services and Applications*, 16(1):419–430. DOI: 10.5753/jisa.2025.5218.
- Pimenta, I. A., Lee, M. H., Bittencourt, L. F., and Gomes, R. L. (2026). Adaptive privacy based on mutual information for machine learning in edge–cloud environments. *IEEE Networking Letters*, 8:49–53. DOI: 10.1109/LNET.2025.3607555.
- Portela, A., Linhares, M. M., Nobre, F. V. J., Menezes, R., Mesquita, M., and Gomes, R. L. (2024). The role of tcp congestion control in the throughput forecasting. In *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*, Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing (LADC 2024), page 196–199, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3697090.3699869.
- Sharma, A., Pandit, S., and Talluri, S. R. (2025). Throughput prediction of fifth-generation cellular system using hybrid feature selection and enhanced sequential decision tree machine learning algorithm. *Wireless Networks*, 31(3):3025–3042. DOI: 10.1007/s11276-025-03917-3.
- Wang, X., Sen, P., Li, R., and Yilmaz, E. (2024). Simulated task oriented dialogues for developing versatile conversational agents. In *Advances in Information Retrieval: 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24–28, 2024, Proceedings, Part I*, page 157–172, Berlin, Heidelberg. Springer-Verlag. DOI: 10.1007/978-3-031-56027-9_10.
- Younas, M. I., Iqbal, M. J., Aziz, A., and Sodhro, A. H. (2023). Toward qos monitoring in iot edge devices driven healthcare—a systematic literature review. *Sensors*, 23(21). DOI: 10.3390/s23218885.
- Zheng, Y., Koh, H. Y., Jin, M., Chi, L., Wang, H., Phan, K. T., Chen, Y.-P. P., Pan, S., and Xiang, W. (2024). Graph spatiotemporal process for multivariate time series anomaly detection with missing values. Available at: <https://arxiv.org/abs/2401.05800>.