Global Conference on Multidisciplinary Research and Innovation
Hosted Online from Berlin, Germany
Date: 2nd February, 2026
Website: https://econferencia.com

MACHINE LEARNING-BASED SELECTION OF INFORMATIVE FEATURES FOR NETWORK TRAFFIC CLASSIFICATION

Tojieva Feruza
Senior Lecturer Department of Infocommunication Engineering Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
11feruza@gmail.com

Abstract
Internet traffic is rapidly increasing as mobile and web applications become more complex and often generate multiple service flows simultaneously, reducing the effectiveness of traditional port-based and DPI approaches. Machine learning enables traffic classification using features derived from packet headers and flow-level statistics; however, not all extracted features contribute equally. Some features improve accuracy, while others introduce noise and computational overhead. Therefore, selecting informative features is crucial for accurate and efficient classification, particularly in near real-time monitoring and traffic management systems. This paper reviews three feature selection families filter, wrapper, and embedded methods and summarizes their main ideas, strengths, and limitations.

Keywords: Machine learning, feature selection, filter, wrapper, embedded.

Today, the volume of Internet network traffic is increasing rapidly. In particular, the growing number of new mobile and web applications makes the traffic composition increasingly complex. Modern applications often generate and transmit multiple service flows simultaneously. This situation reduces the effectiveness of traditional port-based approaches as well as DPI methods.

To address these challenges, the use of artificial intelligence technologies is considered one of the effective solutions. The IP protocol and transport-layer protocols provide many parameters and characteristics derived from packet headers and statistical observations at the flow level. These parameters can be used as initial features for solving network traffic classification problems using machine learning methods[1-3].

However, practical experience shows that not all available features influence the classification process to the same extent. While some features improve model accuracy, others introduce additional computational overhead or even degrade classification performance. Therefore, selecting informative features for network traffic classification is an important scientific and practical task.

For this reason, in network traffic classification, it is first necessary to identify a set of features that is most suitable for solving the problem. The feature selection process is directly related to the key requirements imposed on the classification system accuracy, responsiveness, and computational efficiency. This is especially critical for near real-time network monitoring and traffic management systems, where optimizing the number of features plays a significant role.

Feature selection methods are commonly divided into three main groups: filter, wrapper, and embedded approaches. Filter methods evaluate features using statistical criteria (e.g., information gain, correlation, chi-square (χ^2) test, and others) and perform an initial screening of the feature set.

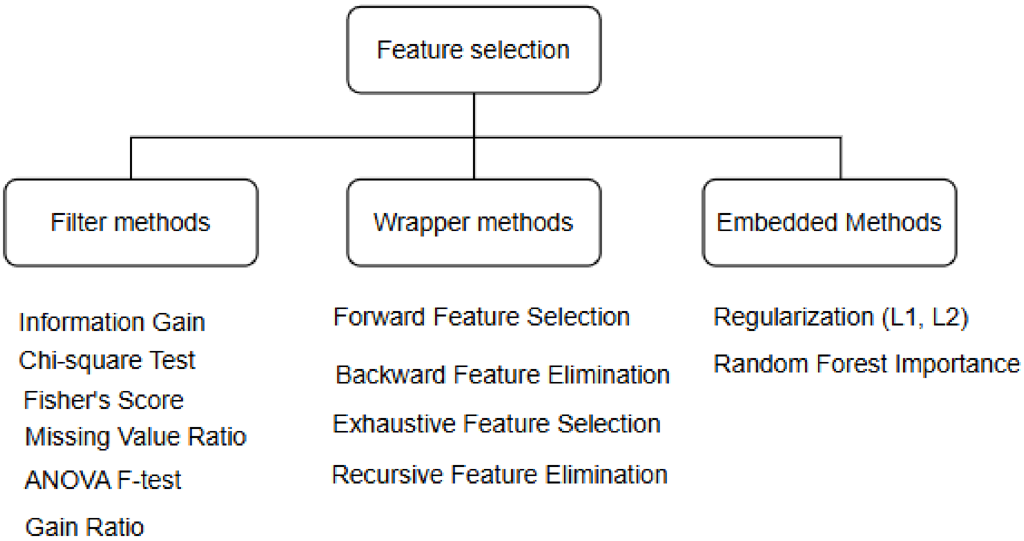

Figure 1. Classification of Feature Selection Methods

These methods are computationally simple and fast, making them convenient for working with large feature sets. However, they do not always fully capture how well the selected features match a specific classifier. In contrast, wrapper methods evaluate a feature subset directly using the performance of a chosen classifier. In other words, different feature combinations are tested based on model accuracy, and the most suitable subset is selected. This approach often provides higher accuracy, but as the number of features increases, computational complexity and processing time grow significantly. Therefore, when applying wrapper methods, it is necessary to consider available computational resources and practical constraints. Embedded methods integrate the feature selection process directly into the model training stage. As a result, feature selection and classifier construction are performed simultaneously. This approach provides a balance between the speed of filter methods and the classifier-specific adaptability of wrapper methods. In particular, embedded methods yield practical and effective results when working with decision trees, Random Forest, and regularization-based models[4].

Table 1. Comparative table of feature selection methods

Method	Main idea	Advantages	Disadvantages
Filter	It evaluates each feature using a statistical criterion.	Very fast; independent of the model; performs well with high-dimensional feature sets.	It does not always capture dependencies between features; some redundant features may remain.
Wrapper	It evaluates feature combinations directly based on the classifiers performance	Often provides higher accuracy; finds a feature subset that is specifically optimized for the chosen model.	Computationally expensive; time-consuming; prone to overfitting.
Embedded	Feature selection is performed within the model training process.	Balances speed and accuracy; practical for real-world use.	Model dependent; the selected features may vary depending on the algorithm.

Thus, in network traffic classification, conducting a comparative analysis of feature selection methods, identifying their strengths and limitations, and choosing the most appropriate approach for a specific task remains a relevant scientific and practical problem.

Filter methods evaluate features independently of the learning model using statistical criteria. Wrapper methods, in contrast, test different feature combinations directly with a selected model and select the subset that yields the best performance. In embedded methods, feature selection is performed as part of the model training process itself.

In many studies, using only one method is often insufficient. Filter methods are computationally efficient but may fail to fully capture dependencies and interactions among features, whereas wrapper methods can improve accuracy more effectively but are computationally expensive. Therefore, modern approaches frequently adopt a two-stage feature selection strategy: in the first stage, the feature space is reduced using a filter method, and in the second stage, the final optimal subset is selected using a wrapper or embedded method[5-6]. As a result of feature selection, the following advantages can be achieved:

- the models accuracy and stability are improved;
- redundant and noisy features are removed;
- training and testing time is reduced;
- model interpretability is enhanced;
- implementation in real-world systems becomes more feasible and practical.

Thus, in machine learning, feature selection is not only a technical optimization step, but also an important methodological process that ensures the model is scientifically justified and practically effective.

CONCLUSION

This paper highlighted the importance of informative feature selection for machine learning-based network traffic classification under modern traffic conditions. A comparative review of filter, wrapper, and embedded methods shows that each approach offers different trade-offs between accuracy, computational cost, and practical deployment. Filter methods provide fast, model-independent screening but may overlook feature interactions. Wrapper methods can achieve higher accuracy by evaluating feature subsets with a classifier, yet they are computationally expensive and may increase the risk of overfitting. Embedded methods perform feature selection during training, offering a balanced and practical solution for real-world systems. Overall, relying on a single method is often insufficient; therefore, a two-stage workflow filter-based reduction followed by wrapper or embedded refinement represents an effective strategy to improve model performance, reduce complexity, and support deployment in operational traffic monitoring and management.

REFERENCES

1. Anjum Nazir, Zulfiqar Memon, Touseef Sadiq, Hameedur Rahman, Inam Ullah Khan, "A Novel Feature-Selection Algorithm in IoT Networks for Intrusion Detection," Sensors, 2023.
2. S. D. Erokhin and A. V. Vanyushina, "Selecting attributes to classify IP traffic by machine learning methods," T-Comm: Telecommunications and Transport, 2018, doi: 10.24411/2072-8735-2018-10140.
3. A. Nazir, Z. Memon, T. Sadiq, H. Rahman, and I. U. Khan, "A Novel Feature-Selection Algorithm in IoT Networks for Intrusion Detection," Sensors, vol. 23, no. 19, Art. no. 8153, 2023, doi: 10.3390/s23198153.
4. Wathiq Laftah Al-Yaseen, Ali Kadhum Idrees, Faezah Hamad Almasoudy, "Wrapper feature selection method based differential evolution and extreme learning machine for intrusion detection system," Pattern Recognition, vol. 132, Art. 108912, 2022, doi:10.1016/j.patcog.2022.108912
5. Abrar Aljehane, Jehad Ali, Fahd N. Al-Wesabi, Manish Kumar, Anwer Mustafa Hilal, Dheya Al-Jumeily, "A novel optimized Artificial Hummingbird Algorithm and deep extreme learning machine for intrusion detection system," Alexandria Engineering Journal, 2024.

6. Yuan Zhou, Guangmin Cheng, Songtao Jiang, Meng Dai, “Building an efficient intrusion detection system based on feature selection and ensemble classifier,” Computer Networks, 2020.