Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

ESTIMATING IMPORTANCE OF HIGHLY CORRELATED FEATURES USING MATRIX FACTORIZATION

Alexander S. Minkin[1],*

[1] *Keldysh Institute of Applied Mathematics*
Moscow,125047, Russia
amink@mail.ru

ABSTRACT

Hyperspectral images contain a large volume of source data that exhibits high correlations along neighboring spectral bands. This makes it necessary to select the most informative features among correlated groups of features to effectively solve various machine learning problems. A method of feature importance evaluation for hyperspectral image data is proposed. This method combines iterative training of Decision Tree classifiers based on spectral features with matrix factorization to overcome sparsity. Decision trees provide intrinsic feature selection mechanism but only a small number of features are usually taken into account by the CART algorithm for training a single decision tree classifier instance. Furthermore when features are highly correlated (e.g., Pearson $\rho > 0.8$), tree-based methods like Random Forest or XGBoost arbitrarily assign importance to one feature while suppressing others, as they redundantly capture the same signal. To overcome this problem, an additional balancing term was incorporated into the optimization function used to obtain the matrix factorization.

The considered method of feature importance evaluation is compared with such model-specific tree-based methods as vanilla Gini impurity decrease and more complicated Boruta algorithm. Classification accuracy is tested using a Random Forest classifier on significant features. Selecting features with higher importance scores yields models boasting greater training accuracy.

Keywords: feature selection, tree-based methods, matrix factorization, machine learning, hyperspectral images

1 INTRODUCTION

Hyperspectral images (HSI) usually acquired from satellites capture light spectra in numerous narrow wavelength ranges. This approach boosts their informational content compared to conventional RGB images. HSI processing applications, including cloud detection, monitoring, agriculture and environmental protection (Borzov and Uzilov, 2016), are challenged by the high dimensionality inherent in large volumes of HSI data. These applications typically rely on machine learning (ML) methods, which can be adversely affected by the curse of dimensionality. A family of techniques designed to overcome the curse of dimentionality are commonly addressed through a set of methods known as dimentionality reduction. In this paper we consider the method for such reduction by selecting a relevant subset of features from both the original and derived sets, without applying any transformations. Furthermore, the high correlations between neighboring spectral bands lead to significant data redundancy. To address this issue and improve the efficiency of subsequent analysis, a special method of feature selection is employed here to identify and retain only the most informative spectral bands and derived features.

The existing methods of feature selection assess both feature properties and target variable relationships, covering forward selection and backward elimination methods, exhaustive search, and ML-based techniques (Guyon and Elisseeff, 2003). Feature importance evaluation can be done directly by analyzing the pair relations between feature and target variable. This approach is generally

*Please send copies to the alternative email address amink@mail.ru

Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

called filter methods. Correlations between adjacent spectral bands gives folks not only to the idea of using dimentionality reduction algorithms for image compression but also to select relevant features (Myasnikov, 2017) from which PCA is the most popular (Zimichev et al., 2014). This approach along with direct analysis of the correlation matrix does not require additional labeling. Such methods as mutual information scores and ANOVA F-value are also very popular filter methods that takes into account labels but doesn't rely on training algorithms. Though these methods are computationally efficient they ignore feature interdependence which can lead to the selection of redundant features that are highly correlated with each other (Zhu et al., 2007). That is why, to select features more effectively, information from the results of HSI classification, in combination with the existing labeling of the source data, is needed. ML models are widely used to evaluate the importance of features, especially in cases of feature correlations.

Model-specific methods for feature selection integrate the feature selection process directly into the model training phase. This approach offers good balance of performance and computational cost (Saito et al., 2018). There are wrapper and embedded model-specific methods. Wrapper methods identify the optimal subset of features by evaluating their different combinations using a specific predictive model. Models such as the recursive feature elimination (RFE) select features iteratively, maximizing the performance of the classification or regression models. Although effective in finding optimal subsets, these methods can be computationally demanding (Zubair et al., 2024). Embedded methods integrate feature selection directly into the model training process itself, using mechanisms like regularization penalties to perform selection simultaneously with parameter estimation. Examples of embedded methods include Lasso (Least Absolute Shrinkage and Selection Operator) and Ridge regression (Fira et al., 2025), neural networks with learnable drop layer (JimÉnez-Navarro et al., 2024), sparse principal component analysis (Seghouane et al., 2019), tree-based methods (Tuv et al., 2009) and so on.

Decision trees provide an intrinsic feature selection mechanism (Breiman et al., 1984), effectively handling non-linear data and correlated features Kohavi and John (1997), and offering inherent explainability Mishra (2022). Decision tree-based models like Random Forests and Gradient Boosting provide feature importance scores based on how much a feature contributes to reducing impurity (e.g., Gini impurity decrease) or variance at each split in the decision trees. More complicated Boruta method determine feature relevance by comparing the importance of an original feature with the importance of permuted counterparts known as shadow features and functions as a wrapper algorithm built around the Random Forest classifier (Kursa and Rudnicki, 2010). As the method tests multiple combinations of original and shadow features to assess feature importance, the process is quite time-consuming.

In general model specific methods give better results than filter methods but their significant drawback is that the selected set of features is inherently tied to the specific ML algorithm being used (Islam et al., 2022). Stochastic nature of some ML models can lead to reproducibility issues, with variations in feature importance rankings across different model training runs (Vos et al., 2024).

Model-agnostic methods offer greater flexibility and are widely used for interpreting complex "black-box" models. Model-agnostic feature selection methods can be applied to any machine learning model, as they are independent of the model's internal workings (Khan et al., 2025). These methods assess the relevance of features based on their intrinsic properties and their relationship with the target variable, without considering a specific predictive model. So they usually require a model to be trained first, but the method itself doesn't care which trainable model is used. Examples of model-agnostic techniques include methods based on eXplainable Artificial Intelligence (XAI) like Permutation Feature Importance (PFI) (Flora et al., 2024) and SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). These techniques offer flexibility and can be used to compare the importance of features across different models but computationally intensive and still affected by feature correlations (Liang et al., 2024) especially PFI (Salih, 2025).

The more sophisticated method of feature importance evaluation is used the more computing time it requires. The more features are taken into account, the more their importance is affected by multicollinearity. In scenarios with a high degree of feature correlation, a Decision Tree classifier offers an efficient way to assess how various feature combinations influence training outcomes, as it operates quickly and is robust to multicollinearity. However for correlated features tree-based methods can assign disproportionate importance to one feature at the expense of others because they redundantly model the same predictive signal. In this case Gini impurity decrease is generally considered

Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

misleading for evaluating feature importance and needs correction. The method of such correction is the topic of the present paper. Multiple evaluations of feature importance across different subsets result in slightly varying and sparse importance values. This paper proposes a method for approximating feature importance values extracted from a sparse matrix, which aggregates results from multiple evaluations on different feature subsets. We suggest the algorithm of feature importance correction by means of matrix factorization obtained via gradient descent process.

2 Problem statement and method of solution

While high-dimensional HSI data offers rich information content, it also presents significant challenges that can degrade the performance of ML models. The aim of this work is to develop an algorithm that enhances model effectiveness by selecting an optimal subset of spectral bands and derived features, thereby improving the prediction accuracy for dense cloud classification.

The iterative training of ML models with different subset of features is used here to approximate target variable. This is done by feature elimination algorithm (Minkin, 2026) by training different instances of Decision Tree classifiers to get multiple values of feature importance. The importance of spectral bands and derived features is determined by assessing how their selection influences the decrease in Gini impurity within Decision Tree classifiers. This approach allows us to rank features according to their contribution to model performance. By evaluating feature subsets during training, we construct a sparse matrix of feature importance scores, with rows corresponding to features and columns to individual trained decision trees. This approach yields the following observations:

1. Minimal subset of features is considered because of correlations between them;
2. Training ML models using as many features as possible is computationally expensive;
3. As a result of multiple training of decision trees, the feature importance matrix is sparse.
4. For correlated features A and B when the tree looks for the best split, it might choose feature A simply because of a tiny, random fluctuation in the data that makes its Gini score marginally better than feature B's at that specific node. Feature A gets credited with the entire impurity decrease for that split. Feature B is ignored at that node because the information it provides is redundant. So the selection becomes arbitrary. In some trees, A is chosen; in others, B is chosen. This leads to unstable feature importance scores. Hence, multiple assessments of feature importance helps reveal their true predictive value.

Logistic regression as linear model are strongly affected by feature correlations (Pourhoseingholi et al., 2008). Decision trees are advantageous here, but training with large feature subsets is still computationally expensive. The resulting importance matrix is sparse, which gives rise to the problem of zero value reconstruction. Such reconstruction is needed to overcome possible feature importance underestimation for correlated features when using Gini impurity decrease as a metric of such evaluation in case of classification problem. When feature importance is approximated from a subset of entries in the sparse matrix, highly correlated features are expected to yield similar importance scores. That process should incorporate additional term during optimization to yield matrix factorization as the result.

So we reformulated the problem by the following way. Feature importance are modeled as the following matrix factorization inspired by recommendation systems key concepts and algorithms from Koren et al. (2009):

$$\hat{T} = PQ^T, \tag{1}$$

where $\hat{T}$ $(n_u \times n_i)$ is the predicted importance corresponding to trained ML model instance u and feature i, P $(n_u \times n_f)$ and Q $(n_i \times n_f)$ are latent factors that capture hidden preferences for features and train instances, respectively. The challenge to the matrix factorization problem is to find P and Q^T. Basically, such an algorithm is going to be used to find latent factors that represent intrinsic feature attributes in a lower dimension (the number n_f of latent factors is chosen beforehand). A learning approach is therefore developed to converge the factorization results close to the observed importance as much as possible, while ensuring all importance values remain nonnegative. Additionally we introduce the feature bias matrix μ as a correction of (1):

$$\hat{T}_{u,i} = \hat{\mu}_{u,i} + \hat{p}_u \hat{q}_i^T. \tag{2}$$

Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

The feature bias matrix μ supposed to capture the tendency of features to have importance higher (or lower) than the average. So μ measures a trained model tendency to systematically overestimate or underestimate feature importance relative to the average across all features and trained ML model instances. Matrices P, Q and μ can be obtained through a regularized optimization procedure:

$$\sum_u \sum_i \left((T_{u,i} - \hat{T}_{u,i})^2 + \lambda(\hat{\mu}_{u,i}^2 + ||\hat{p}_u||^2 + ||\hat{q}_i||^2) \right). \tag{3}$$

Feature importance for correlated features is a key challenge in machine learning interpretability, as standard methods often split or misattribute importance unpredictably. SHAP values become unstable with high multicollinearity (e.g., Pearson $\rho > 0.8$), where one feature might capture all credit due to model fitting order or randomness. PFI can amplify this by perturbing one feature while others compensate via correlation. Correlated features convey redundant information, causing tree-based methods like Gini importance in Random Forest or XGBoost to distribute scores roughly equally among them, masking true group contributions. To overcome randomness in feature importance distribution via tree-based methods the matrix factorization approach is used here with the following additional contribution to the optimization function:

$$\sum_u \sum_i \left((T_{u,i} - \hat{T}_{u,i})^2 + \lambda(\hat{\mu}_{u,i}^2 + ||\hat{p}_u||^2 + ||\hat{q}_i||^2) \right) + E_{u,i,j}, \tag{4}$$

$$E_{u,i,j} = \sum_u \sum_i \sum_j r_{i,j}^2 T_{u,j} (T_{u,j} - \hat{T}_{u,i})^2, \tag{5}$$

where $r_{i,j}$ is Pearson correlation between features i and j. The term $E_{u,i,j}$ in equation (5) penalizes discrepancies in importance scores among highly correlated features, thereby encouraging consistent importance estimates. The penalty term scales positively with both $T_{i,j}$ and the magnitude of the importance discrepancy among correlated features.The derived term takes larger values for features with low importance and smaller values for highly important features.

Stochastic Gradient Descent used here to solve the problem (2) of matrix factorization is an optimization algorithm in which the model parameters (in this case, the bias $\mu_{u,i}$ and factor vectors $\hat{p}_u$ and $\hat{q}_i$) are repeatedly updated by adding the negative of gradients calculated with respect to the function (4) being optimized. The algorithm essentially performs the following steps for a given number of iterations:

$$\begin{aligned}
\hat{\mu}_{u,i} &\leftarrow \hat{\mu}_{u,i} + \gamma \left(\Delta_{u,i} - \lambda \hat{\mu}_{u,i} \right) \\
\hat{p}_u &\leftarrow \hat{p}_u + \gamma \left(\Delta_{u,i} \cdot \hat{q}_i - \lambda \hat{p}_u \right) \\
\hat{q}_i &\leftarrow \hat{q}_i + \gamma \left(\Delta u, i \cdot \hat{p}_u - \lambda \hat{q}_i \right) \\
\Delta_{u,i} &= \delta_{u,i} + \sum_{j \neq i} r_{i,j}^2 T_{u,j} \delta_{u,i,j}
\end{aligned} \tag{6}$$

where γ is the learning rate $\delta_{u,i} = T_{u,i} - \hat{T}_{u,i} = T_{u,i} - (\mu_{u,i} + p_u q_i^T)$ is the error made by the model for the pair (u, i) and $\delta_{u,i,j} = T_{u,j} - \hat{T}_{u,i} = T_{u,j} - (\mu_{u,i} + p_u q_i^T)$ denotes the pairwise distance between the importance score of feature j and the approximated importance score of feature i.

3 Experiments

3.1 Setup

A dedicated set of labeled images of the HYPERION sensor with a spatial resolution of 30 m and a spectral resolution of 10 m in the spectral range of 400–2500 nm was used. After converting the raw radiance data into reflectance values and eliminating zero channels as well as those corresponding to strong light absorption in water vapor, the selected HSI were organised into a training set. The features of this set comprised reflectance values from spectral channels 8–224, combined with derived characteristics and other indices calculated on a pixel-by-pixel basis. Fig. 1 shows the mean spectral reflectance distribution for cloud (green line) and non-cloud (red line) pixels, suggesting that a classification algorithm could be developed to separate these classes based on the spectral characteristics of pixels. In this study, in addition to reflectance values, normalized indices such as NDVI (Huang et al., 2021), NDSI (Jin et al., 2022) and NDWI (Gao, 1996) were also used but they are not shown in Fig. 1.

Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

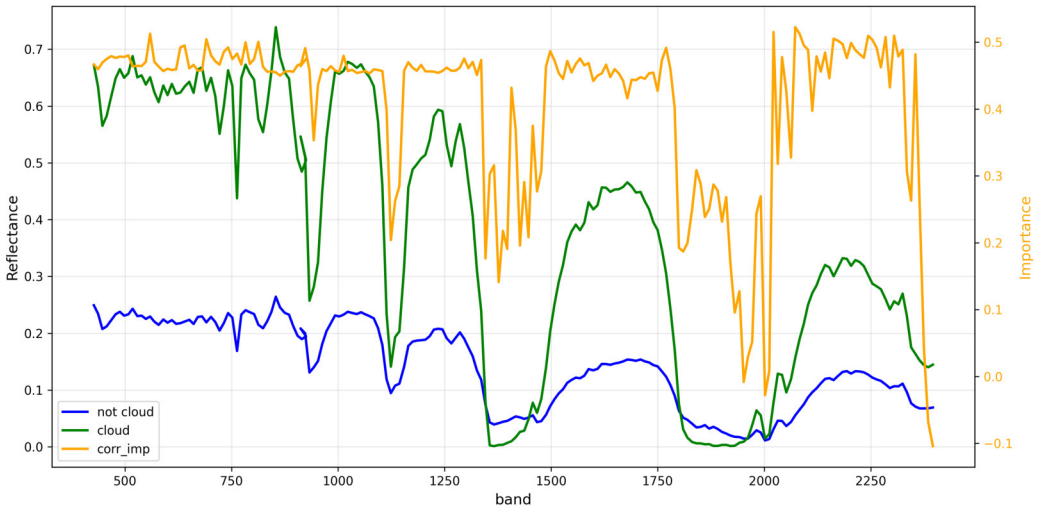

Figure 1: Mean spectral reflectance curves: cloud pixels (green), non-cloud pixels (blue), and correlation feature importance scores (red)

3.2 Results

The input data for the matrix factorization algorithm is feature importance statistics evaluated by training Decision Tree classifier for different values of hyperparameters. For Decision Tree classifier the decrease of Gini impurity is used to assess the feature importance. By considering the accuracy associated with different feature sets, we can calculate the correlation between feature importance and training accuracy and sort features according to this value of correlation. Let's refer to this as correlation feature importance. Fig. 1 shows the effect of assigning greater importance to features with larger differences in spectral reflectance between cloud and non-cloud pixels.

The most important features include NDWI index and the limited set of features from the NIR and the lower SWIR wavelength range. In general there is little to no similarity between the feature importance rankings obtained via the Boruta algorithm and those from the present study, with the exception of the consistently high importance assigned to the NDWI feature and the lower SWIR wavelength range. This can be explained by the use of shadow features in feature analysis via the Boruta algorithm not used here. Fig. 2 presents the random forest model's accuracy as a function

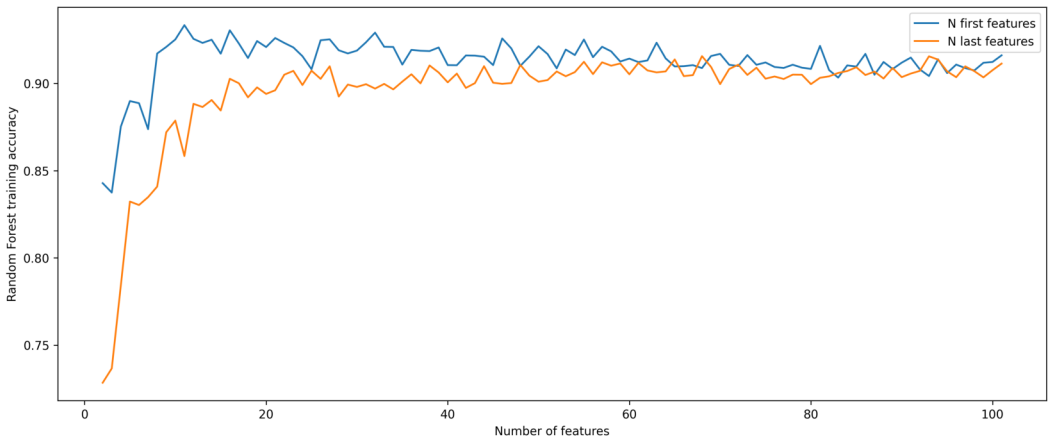

Figure 2: Random Forest training for the first and the last features by their correlation with training accuracy of different Decision Tree classifiers

Mathematics & AI, Vol. 1, No. 2, 2026, Article 1026 DOI: 10.66693/mathai.1026
Special Issue: MathAI 2026 Selected Papers
Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

of the number of selected features. The blue curve illustrates the performance of Random Forest classifier when features are sequentially added following a descending ranking by correlation-based importance (top or first N features). Conversely, the orange curve shows the outcome of adding features in ascending order of importance (bottom or last N features from the same pre-sorted list). Fig. 2 shows that choosing features with higher importance scores (starting from the top of the list in descending order of importance), we obtain models with higher training accuracy.

4 Conclusion

The significant number of channels in HSI, combined with feature multicollinearity, leads to challenges in selecting machine learning models, reducing their accuracy and interpretability. To address this issue for dense cloud classification, decision trees are employed with a selection of a limited number of significant features based on the iterative exclusion algorithm. The proposed method optimizes feature selection by leveraging the correlation between feature importance scores and training accuracy. The most relevant features include NDWI index, limited number of NIR bands and the lower part of SWIR spectrum range bands. Decision Tree model used to assess feature importance effectively handles correlated features avoiding redundancy and can be combined with forward feature selection or backward elimination to achieve robust statistics. The final set of features selected after applying matrix factorization to overcome sparsity with balancing feature importance scores and taking into account the correlation between feature importance and model accuracy. This can be used to construct a classifier for recognizing dense clouds based on their spectral characteristics. Such classifier can be considered a baseline for more complex cloud classification models.

The ambiguity in feature importance rankings produced by traditional methods – such as PFI, SHAP, tree-based methods, and linear model coefficients – underscores the need for a more empirically grounded approach. The proposed strategy of iterative feature elimination with feature importance approximation via matrix factorization offers an alternative solution to this challenge. This approach not only reveals which features are truly critical for maintaining performance but also accounts for feature interactions and helps identify their optimal subset. One of the way to evaluate feature importance is to use Gini impurity decrease in combination with special algorithm for matrix factorization. This study introduces a feature importance evaluation method that combines Gini impurity decrease with a specialized matrix factorization algorithm to approximate the importance of correlated features, refining the final scores using classification accuracy feedback. This was done by rankings derived from the correlation between approximated feature importance scores and classification accuracy to obtain the list of features sorted by their importance.

References

S.M. Borzov and S.B. Uzilov, Borzov, S.M. and Uzilov, S.B. (2016). *J. Comput. Technol.*, 21(1):40 – 48.

(1984). *Classification and Regression Trees*. Chapman and Hall/CRC.

Monica Fira, Liviu Goras, and Hariton-Nicolae Costin, Fira, Monica and Goras, Liviu and Costin, Hariton-Nicolae (2025). Evaluating sparse feature selection methods: A theoretical and empirical perspective. 15(7).

Montgomery L. Flora, Corey K. Potvin, Amy McGovern, and Shawn Handler, Montgomery L. Flora and Corey K. Potvin and Amy McGovern and Shawn Handler (2024). A machine learning explainability tutorial for atmospheric sciences. *Artificial Intelligence for the Earth Systems*, 3(1):e230018. URL `https://journals.ametsoc.org/view/journals/aies/3/1/AIES-D-23-0018.1.xml`.

B.-C. Gao, Gao, B.-C. (1996). Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3):257–266.

Isabelle Guyon and André Elisseeff, Guyon, Isabelle and Elisseeff, André (2003). An introduction of variable and feature selection. *J. Machine Learning Research Special Issue on Variable and Feature Selection*, 3:1157 – 1182.

Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

Sha Huang, Lina Tang, Joseph P. Hupy, Yang Wang, and Guofan Shao, Huang, Sha and Tang, Lina and Hupy, Joseph P. and Wang, Yang and Shao, Guofan (2021). A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1):1–6.

Md Rashedul Islam, Aklima Akter Lima, Sujoy Chandra Das, M. F. Mridha, Akibur Rahman Prodeep, and Yutaka Watanobe, Islam, Md Rashedul and Lima, Aklima Akter and Das, Sujoy Chandra and Mridha, M. F. and Prodeep, Akibur Rahman and Watanobe, Yutaka (2022). A comprehensive survey on the process, methods, evaluation, and challenges of feature selection. *IEEE Access*, 10:99595–99632.

M J JimÉnez-Navarro, M MartÍnez-Ballesteros, I S Brito, F MartÍnez-Álvarez, and G Asencio-CortÉs, JimÉnez-Navarro, M J and MartÍnez-Ballesteros, M and Brito, I S and MartÍnez-Álvarez, F and Asencio-CortÉs, G (2024). Embedded feature selection for neural networks via learnable drop layer. *Logic Journal of the IGPL*, 33(5):jzae062.

Donghyun Jin, Kyeong-Sang Lee, Sungwon Choi, Noh-Hun Seong, Daeseong Jung, Suyoung Sim, Jongho Woo, Uujin Jeon, Yugyeong Byeon, and Kyung-Soo Han, Donghyun Jin and Kyeong-Sang Lee and Sungwon Choi and Noh-Hun Seong and Daeseong Jung and Suyoung Sim and Jongho Woo and Uujin Jeon and Yugyeong Byeon and Kyung-Soo Han (2022). An improvement of snow/cloud discrimination from machine learning using geostationary satellite data. *International Journal of Digital Earth*, 15(1):2355–2375.

Adam Khan, Asad Ali, Jahangir Khan, Fasee Ullah, and Muhammad Faheem, Khan, Adam and Ali, Asad and Khan, Jahangir and Ullah, Fasee and Faheem, Muhammad (2025). Exploring consistent feature selection for software fault prediction: An xai-based model-agnostic approach. *IEEE Access*, 13:75493–75524.

Ron Kohavi and George H. John, Ron Kohavi and George H. John (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1):273–324. URL `https://www.sciencedirect.com/science/article/pii/S000437029700043X`. Relevance.

Yehuda Koren, Robert Bell, and Chris Volinsky, Koren, Yehuda and Bell, Robert and Volinsky, Chris (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.

Miron B. Kursa and Witold R. Rudnicki, Kursa, Miron B. and Rudnicki, Witold R. (2010). Feature selection with the boruta package. *Journal of Statistical Software*, 36(11):1–13.

Annie Liang, Thomas Jemielita, Andy Liaw, Vladimir Svetnik, Lingkang Huang, Richard Baumgartner, and Jason M. Klusowski, Annie Liang and Thomas Jemielita and Andy Liaw and Vladimir Svetnik and Lingkang Huang and Richard Baumgartner and Jason M. Klusowski (2024). Challenges in variable importance ranking under correlation.

Scott Lundberg and Su-In Lee, Lundberg, Scott and Lee, Su-In (2017). A unified approach to interpreting model predictions.

A.S. Minkin, Minkin, A.S. (2026). *Atmospheric and Oceanic Optics*, 39(3):286–295.

(2022). *Practical explainable AI using python: Artificial Intelligence Model Explanations Using Python-Based Libraries, Extensions, and Frameworks*. New York: Apress.

E.V. Myasnikov, Myasnikov, E.V. (2017). Hyperspectral image segmentation using dimensionality reduction and classical segmentation approaches. *Comput. Opt.*, 41(4):564–572.

Mohamad Amin Pourhoseingholi, Yadolah Mehrabi, Hamid Alavi-Majd, and Parvin Yavari, Pourhoseingholi, Mohamad Amin and Mehrabi, Yadolah and Alavi-Majd, Hamid and Yavari, Parvin (2008). Using latent variables in logistic regression to reduce multicollinearity, a case-control example: breast cancer risk factors. *Italian Journal of Public Health*, 5(1).

Shota Saito, Shinichi Shirakawa, and Youhei Akimoto, Saito, Shota and Shirakawa, Shinichi and Akimoto, Youhei (2018). Embedded feature selection using probabilistic model-based optimization. p., 1922–1925. Association for Computing Machinery.

Received: 06 May 2026 · Accepted: 28 May 2026 · Published: 28 May 2026

Ahmed M. Salih, Ahmed M. Salih (2025). Re-visiting explainable ai evaluation metrics to identify the most informative features.

Abd-Krim Seghouane, Navid Shokouhi, and Inge Koch, Seghouane, Abd-Krim and Shokouhi, Navid and Koch, Inge (2019). Sparse principal component analysis with preserved sparsity pattern. *IEEE Transactions on Image Processing*, 28(7):3274–3285.

Eugene Tuv, Alexander Borisov, George Runger, and Kari Torkkola, Tuv, Eugene and Borisov, Alexander and Runger, George and Torkkola, Kari (2009). Feature selection with ensembles, artificial variables, and redundancy elimination. *Journal of Machine Learning Research*, 10:1341–1366.

Gideon Vos, Liza van Eijk, Zoltan Sarnyai, and Mostafa Rahimi Azghadi, Gideon Vos and Liza van Eijk and Zoltan Sarnyai and Mostafa Rahimi Azghadi (2024). Stabilizing machine learning for reproducible and explainable results: A novel validation approach to subject-specific insights.

Zexuan Zhu, Yew-Soon Ong, and Manoranjan Dash, Zhu, Zexuan and Ong, Yew-Soon and Dash, Manoranjan (2007). Wrapper–filter feature selection algorithm using a memetic framework. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(1):70–76.

E.A. Zimichev, N.L. Kazanskiy, and P.G. Serafimovich, Zimichev, E.A. and Kazanskiy, N.L. and Serafimovich, P.G. (2014). Spectral-spatial classification with k-means++ particional clustering. *Comput. Opt.*, 38(2):281–286.

Iqbal Muhammad Zubair, Yung-Seop Lee, and Byunghoon Kim, Zubair, Iqbal Muhammad and Lee, Yung-Seop and Kim, Byunghoon (2024). A new permutation-based method for ranking and selecting group features in multiclass classification. *Applied Sciences*, 14(8).