International Journal of Engineering & Management Informatics
Volume1, Issue01 (2025)

FEATURE ENGINEERING IN MACHINE LEARNING : SELECTION, EXTRACTION, AND THEIR IMPACT ON MODEL PERFORMANCE

ABDULLAH M. NOMAN, AFAQ HUSSAIN1,

Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
Department of CSE

Article Info
Received: 12-03-2025 Revised: 01-05-2025 Accepted: 30-05-2025 Published: 22-06-2025

ABSTRACT

In the data pre-processing stage of machine learning (ML), feature selection and extraction are essential procedures. The models' computational efficiency, interpretability, and performance are all greatly impacted by these stages. Finding the most relevant characteristics in a dataset while maintaining important information is known as feature selection. However, in order to facilitate more efficient learning, feature extraction converts unstructured data into meaningful representations. The present work offers a thorough review of feature selection and extraction, emphasizing its significance, different approaches, and applications in a range of fields. The research also highlights issues like the Curse of Dimensionality, the existence of distracting or noisy characteristics, and the efficacy being limited by the trade-off between interpretability and performance. In order to improve scalability and cross-domain adaptation, the research finishes with insights into new trends and potential paths, such as hybrid techniques and the combination of feature selection with deep learning.

1. Introduction

The quantity of data being gathered and examined using different statistical and machine learning (ML) technologies has increased dramatically during the last several decades. ML is now an essential tool in many fields, including natural language processing, healthcare, education, and finance. The quality and applicability of input characteristics have a significant impact on ML models' effectiveness [1], [2]. High-dimensional, duplicated, or irrelevant data provide challenges for machine learning techniques, which may result in overfitting, increased computing complexity, and other difficulties that impact the model's performance. Feature Engineering (FE) may be used to reduce dimensionality, which can assist address these problems. Feature extraction and feature selection are two of the many phases that make up FE, a method that may reduce high dimensional input data to improve model efficiency and accuracy while maintaining important information. Finding and choosing the most relevant feature subset from the input data that offers the most insightful information while eliminating those that don't significantly aid in the learning process is known as feature selection [1], [3]. In contrast, feature extraction preserves the important information and removes characteristics that aren't essential while transforming the original data into a new, smaller feature set. When raw characteristics are too complicated or inadequate for direct analysis, this is very helpful [3]. When working with high-dimensional data, feature selection and extraction approaches present a number of difficulties despite their many advantages. Sophisticated methods are needed to balance predictive strength and computational efficiency in datasets with hundreds of attributes. Furthermore, it is still difficult to choose or extract the most instructive characteristics without adding bias or omitting important information. The goal of this research is to develop a thorough grasp of feature engineering methods so that scientists may create ML models that are more effective and understandable. Beginning with the introduction, the article offers a thorough review of feature selection and extraction in machine learning. The article is further divided into several parts. After providing the basics in Section 2, Section 3 examines several feature selection methods, and Section

explains methods for feature extraction. The influence of these strategies is discussed in Section 5, and then applications are shown in Section 6. Additionally, section 7 outlines the difficulties as well as possible future paths. The investigation is finally summarized in the conclusion.

2. FundamentalsofFeatureEngineering

3. The act of choosing, altering, or producing new features from unprocessed data in order to enhance the prediction capabilities of an ML model is known as feature engineering. This method is often used after the collection and cleaning of the input data [4]. Feature generation, feature extraction, feature modification, and feature selection are some of the methods used. Effective feature engineering lowers computing costs, enhances generalization, and reduces noise [4]. In machine learning, feature engineering plays the following roles [5, 6]:

- **Enhancing Model Performance:** Models may learn more effectively with high-quality features, which increases accuracy and resilience.
- **Reducing Overfitting:** Models are less likely to overfit when superfluous or unnecessary characteristics are eliminated, which improves generalization.
- **Improving Interpretability:** Models become easier to comprehend and interpret when useful characteristics are chosen, which is especially crucial in fields like healthcare and finance.
- **Optimizing Computational Efficiency:** Model training is accelerated when data dimensionality is decreased.
-

3.1 DifferenceBetween FeatureSelectionand FeatureExtraction

Two main methods in feature engineering that aid in enhancing the caliber of input data are feature selection and feature extraction. But they have distinct functions. While Table 1 shows the distinctions between feature selection and feature extraction based on a few factors, Figures 1 and 2 illustrate the disparities in both procedures.

Table1:DifferencebetweenFeatureSelectionandFeature Extraction

Aspect	FeatureSelection	FeatureExtraction
Definition	Identifies and retains the most relevant features while removing irrelevant or redundant ones.	Transformsrawfeaturesintoanew set of features that better capture essential information.
Approach	Selects from existing features withoutaltering theiroriginal values.	Createsnewfeaturesbycombining or transforming existing ones.
Techniques	Filter methods such as Chi-square, Correlation. Wrapper methods such asRFE,andEmbeddedmethodssuch as LASSO.	PCA, LDA, Singular Value Decomposition (SVD), Autoencoders.
Interpretability	Retainsoriginalfeatures,makingthe model more interpretable.	Can reduce interpretability since transformed features may not have direct real-world meaning.
Use Cases	Suitable when the dataset has redundant or irrelevant features.	Usefulwhenoriginalfeaturesare insufficient or too complex.

FeatureSelectionProcess

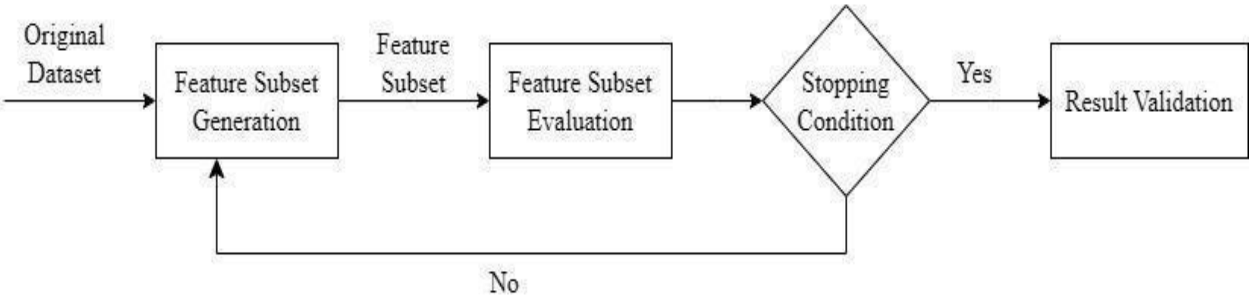

Figure1:FeatureSelectionProcess[7]

FeatureExtractionProcess

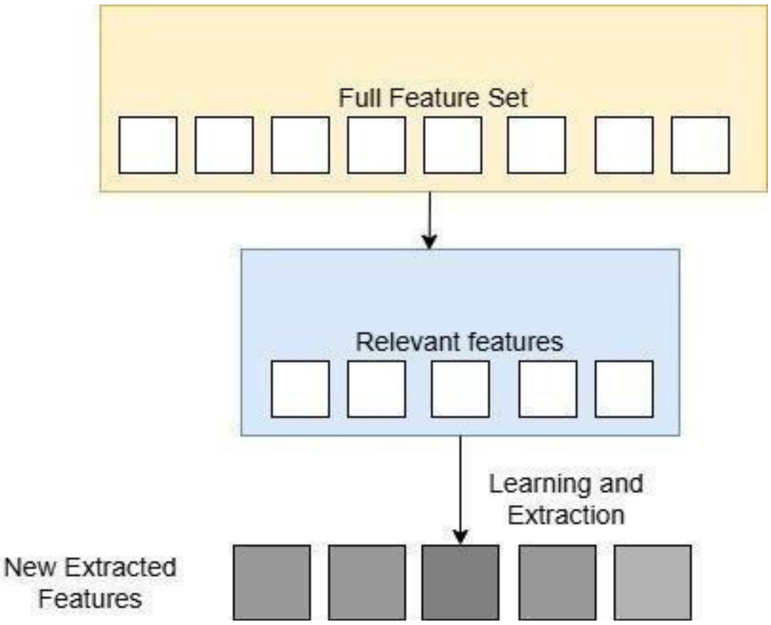

Figure2:FeatureExtractionProcess[7]

3.2 ImportanceofDimensionalityReductioninMachineLearning

A key stage in machine learning is dimensionality reduction, which attempts to minimize features or input variables from high-dimensional datasets without sacrificing important information. Repetitive, noisy, or irrelevant characteristics are often found in high-dimensional data, which might impair model performance [7], [8].Dimensionality reduction has the following main advantages [7], [9]:

1. **Improved Model Performance** – By removing noise from the input information, machine learning models can more easily spot patterns and connections.

   .

2. **Reduced Overfitting** – High-dimensional datasets raise the risk of overfitting, in which the model learns meaningless patterns instead of memorizing noise. Enhancing generality involves eliminating superfluous characteristics.
3. **Faster Computation** – Faster model training and inference result from fewer input variables, which also entail less memory and processing demand.

4. **Better Data Visualization** – Visualizing and analyzing high-dimensional data may be difficult. For improved interpretability, methods like as PCA and t-SNE assist in projecting data into smaller dimensions.
5. **EnhancedInterpretability**–Dimensionality reduction simplifies models by lowering feature complexity, which makes them simpler to interpret and explain.
6. **FeatureSelectionTechnique**s

   In order to enhance the learning performance of classification and prediction models, feature selection seeks to identify a valuable feature subset from a high dimensional dataset [1], [10]. The three main kinds of feature selection approaches are Filter Methods, Wrapper Methods, and Embedded Methods [1], [3], [10], and [11].

3.3 FilterMethods

Filter techniques evaluate an attribute's importance without the use of machine learning algorithms. They rank characteristics according to how well they correlate with the target variable using statistical methods. Filter-based approaches are simple, quick, scalable, and computationally cheap. Some popular filter methods are as follows [1], [11]:

3.3.1 **Correlation-basedFeatureSelection(CFS)**

Pearson's correlation coefficient is the primary tool used in correlation-based feature selection to assess the connection between two variables: the independent variable and the target variable. In order to eliminate duplication, it minimizes inter-feature correlation and chooses characteristics that have a high correlation to the goal since they are more relevant [12]. A better feature subset is indicated by a higher merit score. Text categorization and bioinformatics are two fields that make extensive use of CFS [13], [14]. Additionally, it is limited in its ability to derive non-linear correlations, and if features are highly linked, it may result in the selection of duplicate characteristics.

.

3.3.2 **MutualInformation**

Mutual information calculates both linear and non-linear correlations between two variables by measuring their dependence [12]. Features are seen to be more significant for predicting the target variable if they have higher mutual information scores. Tasks involving image identification and natural language processing benefit greatly from this approach [13], [15].

3.3.3 **Chi-square Test**

To determine if a category characteristic and a target attribute are independent, the chi-square test is used. Finding out whether a characteristic is strongly related to the target variable is done statistically. A significant correlation between the characteristic and the target class is indicated by a high chi-square score. This technique is often used in medical diagnosis and text categorization [9], [13].

3.3.4 **InformationGain**

The decrease in entropy that occurs when a characteristic is employed for categorization is measured by information gain. Its foundation is the idea of entropy, which quantifies the impurity or disorder in a dataset [16]. More relevant aspects are indicated by higher information gain values. Decision tree algorithms like ID3 and C4.5 make extensive use of this method [9].

3.4 WrapperMethods

The assessment criteria for the wrapper techniques is the algorithm's performance, which requires the usage of a predetermined mining algorithm. To improve their overall mining performance, it looks for qualities that are more in line with the mining algorithm. Despite their excellent accuracy, wrapper approaches are computationally costly and sometimes intractable when dealing with large datasets [1], [12], and [13].

3.4.1 RecursiveFeatureElimination(RFE)

It is an iterative process that builds a model using the remaining characteristics after eliminating the least important ones [15], [17], and assesses the model's performance at each stage. Starting with the whole feature set, the procedure progressively eliminates features that contribute less. This process is repeated until a predetermined amount of features is obtained. RFE is often used to improve model performance using machine learning methods like Random Forests and Support Vector Machines (SVM). RFE is especially useful for high-dimensional data since it reduces overfitting and improves model interpretability by removing redundant and unnecessary features. Because RFE requires many iterations of model training and assessment, it may be computationally demanding, particularly when dealing with huge datasets.

3.4.2 ForwardandBackwardFeatureSelection

These are stepwise feature selection techniques that iteratively add or remove

features based on their contribution to model performance[5], [11].

4. By beginning with an empty feature set and gradually adding more important features one at a time, depending on performance improvement, Forward Feature Selection constructs the model incrementally [5, 11, 12]. In contrast, Backward Feature Elimination takes a different strategy. To improve model accuracy, it starts with the whole feature set and then methodically eliminates the least important characteristics one at a time [5, 11]. Backward selection provides a comprehensive evaluation of features, but it is computationally intensive due to the need to train the model many times, particularly for datasets with a large number of features. Evaluation metrics like cross-validation accuracy, AIC (Akaike Information Criterion), or BIC (Bayesian Information Criterion) are used in both approaches to identify features. significance. They are often used when feature interactions are important and computational performance is not the main limitation.

4.1 EmbeddedMethods

tegratingitintotheprocessofmodel training, making them both efficient and

effective[12], [1], [18] thus making them a popular choice for many applications.

4.1.1 ElasticNet(EN)

4.1.2 RidgeRegressionand LASSO(LeastAbsoluteShrinkageandSelection Operator)

Linear models are highly effective for various problems and employ implicit (Ridge) to reduce the number of features and explicit (LASSO) techniques is a useful tool to achieve the shrinkage and variable selection simultaneously[21]. By adding a regularization factor, these models go beyond logistic regression in binary classification. Ridge regression applies regularization by averaging the squared feature weights, while LASSO uses the average of their absolute values[22]. The model's degrees of freedom are limited by regularization constraints, which modify feature weights. In contrast to Ridge, LASSO may efficiently carry out feature selection by reducing certain feature weights to zero. Cross-validation is used to maximize the number of chosen features, which is determined by the degree of regularization [23].

4.2 Hybrid Methods

The efficiency and efficacy of feature selection in machine learning are increased by hybrid feature selection techniques, which combine the benefits of filter and wrapper approaches [24]. Hybrid techniques, which combine the best features of both approaches, are especially helpful for high-dimensional datasets since they strive to strike a compromise between computational efficiency and forecast accuracy. Recent hybrid feature selection research has produced a number of advancements that improve efficacy and efficiency. Even with improvements that include autonomous feature engineering, choosing pertinent features is still crucial. Furthermore, in certain situations, the amount and composition of the data at hand could not allow for the use of intricate models that need on adjusting a number of settings. In recent years, a number of methods have been developed to handle the challenges posed by high-dimensional data.

5. FeatureExtractionTechniques

A crucial stage in machine learning is feature extraction, which turns unprocessed data into a collection of useful characteristics. It is the process of shrinking a primary collection of raw data into new, manageable processing sets while keeping important patterns unchanged. In the words of Brian Ripley [3], "Feature extraction is generally used to mean the construction of linear combinations αTx of continuous features which have good discriminatory power between classes" . These methods may be roughly divided into three categories: deep learning-based, non-linear, and linear [25], [26]. A detailed description of these methods is given in this section [13].

5.1 LinearTechniques

The underlying premise of linear feature extraction algorithms is that data is distributed linearly. These techniques preserve as much variation or class-separability as feasible while projecting data onto a lower-dimensional space [13].

5.1.1 PrincipalComponentAnalysis(PCA)

A linear mapping of the data from a high-dimensional space to a lower-dimensional one, where the variance of the data is raised, is the fundamental idea behind the PCA, an unsupervised approach that reduces dimensionality [3], [27]. In order to reduce the scale dependency, data instances are first gathered by deducting the meaning of each characteristic. The gathered data's covariance matrix is obtained. After that, eigenvectors and eigen values are obtained by performing the eigen decomposition of the covariance matrix [28]. The converted information is defined by the major components, or corresponding eigenvectors, which are chosen based on the biggest eigenvalues. The greatest information diversity is retained by the eigenvectors linked to the higher eigenvalues.

5.1.2 **LinearDiscriminantAnalysis(LDA)**

LDA is a supervised method designed to minimize dimensionality and maximize class separability. Fisher launched it in 1936. It remains one of the well-organized statistically based pattern categorization methods. LDA takes class labels into account and determines a linear projection that optimizes the ratio of between-class variance to within-class variation, in contrast to PCA, which concentrates on variance. In the feature space, this method detects a projection vector to raise the between-class scatter matrix and lower the within-class scatter matrix [5, 10].

5.2 Non-Linear Techniques

When the underlying data distribution is non-linear and sophisticated modifications are needed to capture key patterns, non-linear feature extraction techniques are used. Nonlinear techniques may capture complex connections between input characteristics and output variables without needing domain knowledge or previous preconceptions about the data, often improving forecast accuracy [26].

5.2.1 **KernelPCA**

By using the kernel approach, which translates data into a higher-dimensional space where linear separation is feasible, Kernel PCA expands on PCA [27]. It is especially helpful for jobs with elaborate patterns since it makes it possible to capture complex structures and non-linear interactions between variables [29], [30], and [31].

5.2.2 **IndependentComponentAnalysis(ICA)**

A statistical technique called ICA is used to distinguish distinct sources from mixed inputs. It is often used to blind source separation issues like brain imaging data analysis and mixed audio signal separation. ICA is helpful for extracting significant independent features since it makes the assumption that the underlying signals are non-Gaussian and statistically independent. Reducing the statistical reliance of component illustration is the primary proof of this linear conversion approach [3]. Solutions in neuroscience that support the same objective of reducing redundancy and elucidating some aspects of the brain's initial processing of sensory information promote the widespread use of ICA for FE.

5.3 DeepLearning-basedFeatureExtraction

As deep learning has grown in popularity, feature extraction methods have expanded to include methods that automatically pick up hierarchical data representations.

5.3.1 **Autoencoders**

Neural Networks (NN) designed for unsupervised feature learning are autoencoders. For feature learning, they use the back propagation method [26]. They consist of a decoder that reconstructs the original input and an encoder that compresses input data into a lower-dimensional representation. The learnt latent representation is often used for anomaly detection and dimensionality reduction, and it captures key patterns [13].

5.3.2 ConvolutionalNeuralNetworks(CNNs)forFeature Learning

CNNs are often used for feature extraction in computer vision applications. Convolutional layers are useful for recognizing edges, textures, and object components because they naturally learn the spatial hierarchies of features [32]. For feature extraction in transfer learning settings, pre-trained CNN models like VGG, ResNet, and EfficientNet are often In machine learning, feature extraction is still an essential phase that allows for effective data representation and enhances model performance. The kind of data and the particular needs of the work at hand determine which extraction method is best [8].

6. ImpactofFeatureSelectionandFeature Extraction

Both feature extraction and feature selection are useful methods for improving learning performance, especially when it comes to accuracy, model interpretability, computational efficiency, and storage requirements. Because feature selection is more readable and interpretable than the other option, it is often chosen. Maintaining unique characteristics in the chosen subset is very beneficial in a variety of research fields, such as identifying pertinent genes connected to certain illnesses in the medical sector [5]. The primary importance of feature selection techniques is in their ability to speed up training for machine learning techniques. These methods make the models less complicated and make understanding easier. On the other hand, the feature extraction approach is crucial for lowering processing resource requirements without losing crucial data. The amount of duplicated data for analysis is decreased by this technique [26].

Accuracy

• By eliminating duplicate and noisy features, feature selection improves accuracy and creates a cleaner dataset. • By producing more representative features that reflect intricate connections within the data, feature extraction may further increase accuracy.
• To maximize performance, feature selection and extraction may sometimes be combined.

ComputationalEfficiency

• Because feature selection speeds up model convergence by using fewer features during raining, it often lowers computing costs.• Feature extraction may increase efficiency when reducing high-dimensional input spaces, but it can also be computationally demanding, especially when using sophisticated approaches like deep learning.

Interpretability

• By keeping the original feature names and values, feature selection preserves interpretability, which facilitates comprehension and justification of model choices. • Because modified features often don't directly correlate with original properties, feature extraction drastically limits interpretability, which makes explanation-based AI approaches more difficult.

7. ApplicationsforFeatureSelectionandExtractionin Machine Learning

Below are some key applications of feature selection and extraction in different fields[13], [34]:

7.1 Healthcare

In healthcare, feature selection and extraction are essential for developing accurate and efficient predictive models. Some key applications include:

- **Disease Prediction:** using genetic data to identify pertinent biomarkers for the diagnosis of conditions including diabetes, cancer, and heart disease.
- **MedicalImageAnalysis:**identifying cancers, fractures, and anomalies by identifying key characteristics in X-ray, MRI, and CT scan pictures.

7.2 **Electronic Health Records (EHRs):** Reducingduplicate patient data while preserving crucial elements for tailored treatment recommendations and predictive analytics.

7.3 Finance

Thefinancesectorleveragesfeatureselectionandextractiontoenhancesecurityan drisk assessment models. Applications include:

- **Fraud Detection:** Finding unusual transactions via picking out important transactional patterns that point to fraud.
- **Credit Risk Analysis:** identifying key elements in financial histories to evaluate a person's or company's creditworthiness.
- **Stock Market Prediction:**choosing key market indicators that have an effect on stock prices and making sure forecasting models are reliable.

7.4 NaturalLanguage Processing

InNaturalLanguageProcessing(NLP),featureselectionandextractioncontribute significantly to text-based machine learning models[26]. Applications include:

- **TextClassification:**determining which words, n-grams, or syntactic structures are most relevant for classifying documents [18].
- **SentimentAnalysis:**identifying the emotion of social media messages, comments, and customer reviews by extracting sentiment-bearing words and phrases

.

7.5 ComputerVision

Featureselectionandextractionarecriticalinprocessingvisualdataforvariouscomputer vision applications:

- **ImageRecognition:**choosing discriminative characteristics to distinguish faces, objects, or patterns in pictures.
- **Object Detection:** identifying things in photos and movies by extracting important characteristics including edges, textures, and forms.

7.6 **Facial Recognition:** enhancing recognition accuracy by using feature extraction methods like deep learning-based embeddings or Principal Component Analysis (PCA).

7.7 Cybersecurity

In cybersecurity, feature selection and extraction enhance the performance of security- related machine learning models. Key applications include:

- **Intrusion Detection:** identifying anomalous activity and security concerns by choosing the most relevant network traffic aspects [18], [35].
- **MalwareAnalysis:**identifying and categorizing malware by extracting characteristics from executable files and network traffic patterns.
- **User Authentication:** enhancing access control and identity verification via the use of behavioral and biometric feature extraction techniques. Effective use of feature selection and extraction results in machine learning models that are more accurate, compute more quickly, and are easier to understand, all of which enhance decision-making and practical applications.
- **ChallengesandFuture Directions**
- Despite their benefits, feature selection and extraction have a number of drawbacks that restrict their use in certain fields. The Curse of Dimensionality, which occurs when high-dimensional datasets increase computational complexity and the danger of overfitting, is one of the main concerns that many researchers have explored. This makes it challenging to find the most useful features [5, 17, 12, 18]. This problem is exacerbated by noisy and irrelevant features, which, if improperly handled, may mislead learning algorithms and impair model performance [35], [12]. Challenges with scalability and computational complexity also arise since conventional methods often fail when used on real-time or large-scale datasets [5, 17].

In order to handle large amounts of data efficiently, scalable algorithms are required. Furthermore, whereas techniques like PCA and autoencoders might increase accuracy, they often fall short in terms of interpretability and explainability [5]. Because skewed datasets may lead to biased feature prioritization, which limits generalizability, data imbalance and bias can have an influence on the dependability of feature selection [36]. Another challenge is domain dependency, when techniques that work well in one area may not work well in another. Moreover, integration with deep learning is still in its infancy. Automatic feature learning is usually the foundation of deep learning models, which results in hybrid techniques that specifically include feature selection. To tackle these obstacles, creative and multifaceted research approaches are needed:

- • Feature selection performance may be streamlined and enhanced by hybrid and automated techniques that include filter, wrapper, and embedding approaches, particularly using AutoML [36], [37]. • Resilience and equity may be increased by developing robust techniques for noisy and unbalanced data, such as robust feature weighting and synthetic data creation. The development of dimensionality reduction strategies that work with neural network designs may be aided by ongoing research in feature selection for deep learning [36]. Improving generalization across a range of applications may be achieved by enhancing cross-domain adaptability via transfer learning and domain adaptation strategies.

8. Conclusion

A key component of machine learning, feature engineering is still essential for improving the generalizability, accuracy, and efficiency of models. High-dimensional datasets provide considerable difficulties, however, since they often include duplicated or unnecessary characteristics that may confuse learning algorithms and impair performance. In order to reduce dimensionality, mitigate overfitting, and enhance interpretability, this study looked at both feature selection and feature extraction strategies. The article briefly discusses both linear and non-linear feature extraction techniques as well as a variety of feature selection strategies, including filter, wrapper, embedded, and hybrid approaches. Issues like scalability, interpretability, and domain adaptability still exist despite the advancements in this sector. It is anticipated that future developments in explainability, automation, and integration with deep learning architectures will spur innovation and improve the efficacy of machine learning models in a variety of real-world applications.

References

[1] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” Computers and Electrical Engineering, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[2] K.HopfandS.Reifenrath,“FilterMethodsforFeatureSelectioninSupervisedMachine Learning Applications -- Review and Benchmark,” Nov. 2021, [Online]. Available: http://arxiv.org/abs/2111.12140

[3] S. Khalid, “A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning,” 2014. [Online]. Available: www.conference.thesai.org

[4] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. vandenBroucke, “Special issue on feature engineering editorial,” Mach Learn, vol. 113, no. 7, pp. 3917–3928, Jul. 2024, doi: 10.1007/s10994-021-06042-2.

[5] A.Ullah, F. H.Khan,U. Qamar, andS.Bashir,“Dimensionalityreduction approaches and evolving challenges in high dimensional data,” ACM International Conference Proceeding Series, pp. 1–8, 2017, doi: 10.1145/3109761.3158407.

[6] E. Buraimoh, R. Ajoodha, and K. Padayachee, “Importance of Data Re-Sampling and Dimensionality Reduction in Predicting Students’ Success,” 3rd International Conference on Electrical, Communication and Computer Engineering, ICECCE 2021, no. June, pp. 12–13, 2021, doi: 10.1109/ICECCE52056.2021.9514123.

[7] S. Velliangiri, S. Alagumuthukrishnan, and S. I. Thankumar Joseph, “A Review of Dimensionality Reduction Techniques for Efficient Computation,” in Procedia Computer Science, Elsevier B.V., 2019, pp. 104–111. doi: 10.1016/j.procs.2020.01.079.

[8] H. Motoda and H. Liu, “Feature selection, extraction and construction,” Communication of IICM, vol. 5, pp. 67–72, 2002.

[9] J. Novaković, P. Strbac, and D. Bulatović, “Toward optimal feature selection using ranking methods and classification algorithms,” Yugoslav Journal of Operations Research, vol. 21, no. 1, pp. 119–135, 2011, doi: 10.2298/YJOR1101119N.

[10] F. Saberi-Movahed, K. Berahman, R. Sheikhpour, Y. Li, and S. Pan, “Nonnegative Matrix Factorization in Dimensionality Reduction: A Survey,” May 2024, [Online]. Available: http://arxiv.org/abs/2405.03615

[11] I. Guyon and A. M. De, “An Introduction to Variable and Feature Selection André Elisseeff,” 2003.

[12] X. Cheng, "A Comprehensive Study of Feature Selection Techniques in Machine LearningModels,"InsightsinComputer,SignalsandSystems,vol.1,no.1,Nov.2024, [Online]. Available: https://soapubs.com/index.php/ICSS

[13] B. Ghojogh et al., "Feature Selection and Feature Extraction in Pattern Analysis: A Literature Review," May 2019, [Online]. Available: http://arxiv.org/abs/1905.02845

[14] B. Kaur, A. Gupta, and R. K. Singla, "Predicting Student Performance in Academics using Machine Learning Approach: A Case Study," 3rd International conference on Latest Trends in Engineering and Management (ICLTEM), pp. 1–12, 2021.

[15] B. Venkatesh and J. Anuradha, A Hybrid Feature Selection Approach for Handling a High-Dimensional Data, vol. 74, no. January. Springer Singapore, 2019. doi: 10.1007/978-981-13-7082-3_42.

[16] L.Chuang,C.Ke,andC. Yang,"A HybridBothFilterandWrapper FeatureSelection Method for Microarray Classification," 2008, pp. 146–150.

[17] K.YanandD.Zhang,"Featureselectionandanalysisoncorrelatedgassensordata with recursive feature elimination," Sens Actuators B Chem, vol. 212, pp. 353–363, 2015, doi: https://doi.org/10.1016/j.snb.2015.02.025.

[18] V.KumarandS.Minz,"FeatureSelection:AliteratureReview,"TheSmartComputing Review, vol. 4, no. 3, pp. 211–229, 2014, doi: 10.6029/smartcr.2014.03.007.

[19] F. Amini and G. Hu, "A two-layer feature selection method using Genetic Algorithm and Elastic Net," Expert Syst Appl, vol. 166, Mar. 2021, doi: 10.1016/j.eswa.2020.114072.

[20] S. Ghosh and M. Alamgir Islam, "Performance Evaluation and Comparison of Heart Disease Prediction Using Machine Learning Methods with Elastic Net Feature Selection," Am J Appl Math Stat, vol. 11, no. 2, pp. 35–49, Apr. 2023, doi: 10.12691/ajams-11-2-1.

[21] Y.KimandJ.Kim,“GradientLASSOforfeatureselection.”

[22] MuthukrishnanRandRohiniR,“LASSO:AFeatureSelectionTechniqueInPredictive ModelingForMachineLearning,”in2016IEEEInternationalConferenceonAdvances in Computer Applications (ICACA), IEEE, 2016.

[23] Prabha M. Kumarage, B. Yogarajah, and Nagulan Ratnarajah, “Efficient Feature Selection for Prediction of Diabetic Using LASSO,” in 2019 19th International Conference on Advances in ICT for Emerging Regions (ICTer), Institute of Electrical and Electronics Engineers (IEEE), 2019.

[24] H.H.Hsu,C. W. Hsieh, and M. Da Lu,“Hybridfeatureselection by combining filters and wrappers,” Expert Syst Appl, vol. 38, no. 7, pp. 8144–8150, 2011, doi: 10.1016/j.eswa.2010.12.156.

[25] Ö. Uncu and I. B. Türkşen, “A novel feature selection approach: Combining feature wrappers and filters,” Inf Sci (N Y), vol. 177, no. 2, pp. 449–466, 2007, doi: 10.1016/j.ins.2006.03.022.

[26] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, “Autoencoders and their applications in machinelearning: a survey,” ArtifIntell Rev, vol. 57, no. 2, Feb. 2024, doi: 10.1007/s10462-023-10662-6.

[27] Foo Hui Mean and Yuan-chin Ivan Chang, “A Survey: Potential Dimensionality Reduction Methods For Data Reduction,” Feb. 2025, [Online]. Available: http://arxiv.org/abs/2502.11036

[28] L. C. Borges, V. M. Marques, and J. Bernardino, “Comparison of data mining techniques and tools for data classification,” Proceedings of the International C* ConferenceonComputer Science andSoftwareEngineering -C3S2E’13, no.October 2014, p. 113, 2013, doi: 10.1145/2494444.2494451.

[29] R. M. Ebied, “Extraction using PCA and Kernel-PCA for Face Recognition,” in th International Conference on INFOrmatics and Systems (INFOS2012) - 14-16 May Computational Intelligence and Multimedia Computing Track, May 2012.

[30] R. Rosipal, M. Girolami, L. J. Trejo, and A. Cichocki, "Kernel PCA for Feature Extraction and De-Noising in Nonlinear Regression," Neural Comput&Applic, vol. 10, pp. 231–243, 2001.

[31] A. M. Jade, B. Srikanth, V. K. Jayaraman, B. D. Kulkarni, J. P. Jog, and L. Priya, "Feature extraction and denoising using kernel PCA," Chem Eng Sci, vol. 58, no. 19, pp.4441–4448, 2003,doi: 10.1016/S0009-2509(03)00340-3.

[32] Suresh Dara and Priyanka Tumma, "Feature Extraction By Using Deep Learning: A Survey," in Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018), Coimbatore, India: IEEE, Mar. 2018. doi: 10.1109/ICECA.2018.8474912.

[33] N. Bosch and L. Paquette, "Unsupervised Deep Autoencoders for Feature Extraction with Educational Data." [Online]. Available: https://github.com/pnb/dlwed17

[34] D. Ruano-Ordás, "Machine Learning-Based Feature Extraction and Selection," Aug. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/app14156567.

[35] A. Sivasubramanian, M. Devisetty, and P. Bhavukam, "Feature Extraction and Anomaly Detection Using Different Autoencoders for Modeling Intrusion Detection Systems," Arab J Sci Eng, vol. 49, no. 9, pp. 13061–13073, Sep. 2024, doi: 10.1007/s13369-024-08951-5.