**Abstract**

From early times, humanity has expressed a profound need to predict the future. In ancient Greece, oracles held a place of high respect, and influence. In modern times, the quest for accurate forecasting, has shifted to the realms, of research and science∙ empowered by advanced computational tools, provided by Artificial Intelligence, particularly Machine Learning. One of the fields, where Machine Learning, has established a fertile ground is in the domain, of forest fire management.

Forest fires, pose a major threat to both human, and animal life, with significant economic and social impacts. A reliable prediction system is crucial for mitigating these effects, especially during summer months, dry seasons, and in high-risk areas, such as the Mediterranean.

This study, explores feature construction, and selection methods, applied to forest fire data, collected over 10 years in Greece, incorporating prevailing weather conditions at ignition, and during suppression. By applying techniques like Principal Component Analysis (PCA), Minimum Redundancy Maximum Relevance (MRMR) feature selection, and Grammatical Evolution for feature construction, this research aims to identify key factors influencing fire duration.

These techniques, have become invaluable allies, in addressing complex predictive challenges, and advancing, our understanding of future events.

Our approach, leverages advanced computational methods, to analyze complex datasets, providing a deeper understanding of the primary drivers, of wildfire behavior, and enabling more effective mitigation strategies.

**Feature Construction**

To improve the predictive accuracy of Machine Learning (ML) models in forest fire analysis, there is increasing emphasis on leveraging **Feature Construction techniques**. These methods involve creating new, meaningful variables by combining or transforming existing data attributes. For example, integrating material resources deployed during a forest fire event into a single metric constitutes feature construction, enabling models, to better capture the complexity, of fire incidents, and resource allocation. Another example, for Feature Construction, during a forest fire, is combining weather attributes, in order to form, a fire risk index. Such, approaches enhance data representation, facilitating more robust, and interpretable predictive models, in disaster management.

Τhis paper provides a concise overview of Feature Construction methods, applied to predicting forest fires, and improving early detection, systems. It also offers a comprehensive review of current models used in this field, highlighting their strengths and limitations.

In the domain of fire management, early warning systems for fire outbreaks serve as the cornerstone for timely containment and mitigation efforts. Over recent years, extensive research has been conducted to enhance fire detection methodologies through advanced computational techniques.

For instance, [Harkat et al. (2022)](#_References.) utilized **feature construction** to train a model on multidimensional data. By eliminating irrelevant or redundant features and selecting the most relevant ones for classification, their approach achieved a remarkable accuracy of **96.21%.** However, their study also revealed limitations in the performance of deep learning (DL) models under constrained conditions.

Another significant contribution is the work by [Zhang et al. (2021),](#_References.) which demonstrated that every feature is integral to the overall performance of their algorithm. The combined use of features resulted in a highly impressive accuracy rate of **99.02%** for forest fire recognition. Nevertheless, the study identified a critical weakness: the removal of a single feature (mean) led to a substantial increase in false positives, underscoring its importance in minimizing errors.

Similarly, [Yang et al. (2020)](#_References.) explored the construction of novel, dynamic features to improve accuracy and reduce false alarms in fire detection systems. By combining smoke and flame characteristics, their approach focused on enhancing early warning capabilities. However, the study had notable limitations, as it failed to disclose the source of their data or the specific results obtained, thereby impacting its reproducibility and transparency.

At this point, we will mention a novel method, that was introduced by Tsoulos, for Featured Constructions. In order to improve the effectiveness of Artificial Intelligence, tools, such as Radial basis function (RBF), Multi-layer perceptron (MLP), K-nearest neighbor (KNN), and Neural Networks. The proposed method, is an experimental comparison, carried out against the accuracy obtained, on the original features, as well on features created by the PCA method, [Tsoulos et al. 2008.](#_References.)

**Principal Component Analysis (PCA)**

The **Principal Component Analysis** (PCA) technique, introduced by mathematician Karl Pearson in 1901, and developed by Harold Hotelling (1933). This technique, operates on the principle when data from a higher-dimensional space is transformed into a lower-dimensional space, the resulting lower-dimensional representation should retain the maximum variance, of the original data.

Notably, it is worth mentioning that the use of PCA, on larger datasets, became practical only after the advent of electronic computers, which made it computationally feasible, to handle datasets, beyond trivial sizes [(Cadima & Jolliffe, 2016).](#_References.)

Continuing, with the applications of PCA, it is a widely utilized technique in exploratory data analysis, and machine learning∙ particularly, in building predictive models. It is an unsupervised learning method, designed to analyze, the relationships among a set of variables. Often referred to as a form of general factor analysis, it involves regression to determine a line of best fit. The primary objective, of PCA, is to reduce the dimensionality of a dataset, while retaining the most significant patterns, and relationships, among the variables, all without requiring prior knowledge, of the target variables [(i2tutorials, 2019).](#_References.)

Next, we will briefly reference studies, that have utilized PCA, covering different areas, such as: statistical physics, genetic improvement, face recognition, economic & environmental sciences, medical prediction, e.t.c.

Explicitly, the research conducted, by Park (2024), highlights the reasons behind the success of the PCA technique, for lattice systems. The study's primary limitation lies in the dependency of the proposed formula's accuracy on the dataset size. Specifically, the results achieve full precision, only under the condition of an infinite dataset. This constraint restricts the practical applicability, of the method, when working with finite or limited data, a common scenario, in real-world analyses, [Park (2024).](#_References.)

The following research, by Sarma, (2024), performed with PCA, in order to evaluate, morphometric traits, under a multivariate approach. The findings suggest, that PCA could significantly enhance the genetic improvement, [Sarma et al. (2024).](#_References.) Noteworthy, is the fact, that the 64.29%, of the total variance explained, can be considered relatively low. This suggests, that a significant amount of unexplained information remains, which is not captured, by the four principal components.

The next study, by Parente, the PCA was used for monitoring the cultivation, of cannabis, in Albania. Specifically, with PCA they remove redundant spectral information from multiband datasets, [Parante et al. (2021).](#_References.)

The article, by [Slavkovic & Jevtic (2012),](#_References.) presents the implementation of a face recognition system based on the Principal Component Analysis (PCA) algorithm.

The PCA technique was utilized, by Hargreaves, for stock selection, specifically to identify, a limited number of stock variables, that could effectively aid, in determining winning stocks, [Hargreaves & Mani (2015).](#_References.)

The following documents present a modified model of Principal Component Analysis (PCA).

The paper, by Xu (2024), presents an interesting example of a modified application of Principal Component Analysis (PCA)∙ utilizing, both linear and non-linear methods, through Kernel PCA (KPCA), in combination, with the Adaptive Boosting (AdaBoost) algorithm, [Xu (2024).](#_References.)

The study, by Zhang (2022), a neural network model, combining PCA and Levenberg-Marquardt Backpropagation (LMBP) was developed, to efficiently, and accurately analyze, and predict the interaction between IAQ and its influencing factors. In particular, it was examined indoor air quality (IAQ), and its relationship, with building features, and environmental conditions, [Zhang (2022).](#_References.)

In the next research, by Akinnuwesi, it was employed a hybrid approach, combining Principal Component Analysis (PCA), and Support Vector Machine (SVM). They create, the Breast Cancer Risk Assessment and Early Diagnosis (BC-RAED) model, designed to accurately detect BCa, in its early stages. PCA, was initially applied to extract features, during the first preprocessing stage, followed by further feature reduction, in the second stage. The multi-preprocessed data, were analyzed for breast cancer risk, and diagnosis using SVM. The BC-RAED model, achieved, an accuracy of 97.62%, a sensitivity of 95.24%, and a specificity of 100% in assessing, and diagnosing breast cancer risk, [Akinnuwesi et al. (2020).](#_References.)

Subsequently, we will briefly mention certain studies, that have been conducted, in the field: of Forest Fires.

Guan's, research focuses on forest fire prediction using PCA-preprocessed data. The preprocessing step removed irrelevant information, simplifying analysis. Linear regression and random forest methods were then applied, revealing temperature, relative humidity, wind, and rain as the most influential factors in forest fire occurrence, [Guan (2023).](#_References.)

A novel model was developed, by Nikolov, using meteorological forecast data as input. Principal Component Analysis (PCA), with orthogonal rotation, was applied to reduce 195 meteorological variables, from the NARR dataset, to a smaller set of significant fire-ignition predictors, later used in logistic regression, to calculate wildfire ignition probabilities, [Nikolov et al. 2022](#_References.).

Like the aforementioned study, this one also belongs to Nikolov. This research, focuses on predicting wildfire ignitions, caused by lightning strikes, which account for the largest area burned annually, in the extratropical Northern Hemisphere. Principal Component Analysis (PCA), played a key role, in reducing 611 potential predictors, to 13 principal components, which were used in logistic regression to identify the primary factors influencing lightning occurrence, [Nikolov et al. 2024.](#_References.)

**Minimum Redundancy Maximum Relevance (MRMR)**

The min-redundancy max-relevance algorithm, introduced by Chris Ding and Hanchuan Peng, in their 2005 paper titled: “Minimum Redundancy Feature Selection from Microarray Gene Expression Data.” The (MRMR) aims to optimize feature selection, by minimizing redundancy, and maximizing relevance by [Ramirez – Gallego et all. (2017).](#_References.)

In sum, MRMR enhances relevance-only methods, such as using an f-test between the target, and the features. When two features are similar, MRMR prioritizes only the one, with the highest relevance.

Στην συνέχεια, θα εξετάσω, αυτές τις πηγές, για mrmr.

<https://ar5iv.labs.arxiv.org/html/1908.05376>

<https://ieeexplore.ieee.org/document/5136466>

<https://www.researchgate.net/publication/372959977_Air_quality_prediction_model_based_on_mRMR-RF_feature_selection_and_ISSA-LSTM>

<https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1419316/full>

<https://www.mdpi.com/2076-3417/10/7/2255>

<https://ieeexplore.ieee.org/document/10643969>

# 

# **References.**

**1)** [Harkat, H., Nascimento, J. M.P., Bernardino, A., Ahmed, H.F.T. *Fires Images Classification based on a handcraft approach.* Expert Systems with Applications, volume 212, February 2023, 118594. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0957417422016499 (accessed on November 16, 2024).](#A1)

2) [Zhang, C., Wu, D., Ji, L., Ran, R., Wu, H., Xu, Y.](#A2) *[Forest Fire Recognition Based on Feature Extraction from Multi – View Images](#A2)*[. Traitement du Signal, volume 38, No 3, pp. 775 – 783. June, 2021. Available online: https://www.researchgate.net/publication/353490623\_Forest\_Fire\_Recognition\_Based\_on\_Feature\_Extraction\_from\_Multi-View\_Images (accessed on 16 November 2024).](#A2)

3) [Yang, Y., Sun, F., Lin, C., Liu, Z., Cho, L.](#A3) *[Forest Fire Compound Feature Monitoring Technology Based on Infrared and Visible Binocular Vision](#A3)*[. Journal of Physics: Conference Series, 1792 (2021) 012022. Available online: https://iopscience.iop.org/article/10.1088/1742-6596/1792/1/012022/pdf (accessed on 16 November 2024).](#A3)

4[) Cadima, J., Jolliffe, I.T.](#A4) *[Principal component analysis: a review and recent developments.](#A4)* [National Library of Medicine. 2016. Available online: https://pmc.ncbi.nlm.nih.gov/articles/PMC4792409/ (accessed on 16 November 2024).](#A4)

5) [i2tutorials. *What are the Pros and Cons of the PCA*? October 1st, 2019. Available online: https://www.i2tutorials.com/what-are-the-pros-and-cons-of-the-pca/ (accessed on 16 November 2024](#A5)).

6) [Ramirez – Gallego, S., Lastra, I., Martinez – Rego, D., Bolon – Canedo, V., Benitez, J.M., Herrera, F., Alonso – Betanzos, A. *Fast-mRMR: Fast Minimum Redundancy Maximum Relevance Algorithm for High – Dimensional Big Data.* International Journal of Intelligent Systems, volume 32, pp. 134 – 152. 2017. Available online: https://sci2s.ugr.es/sites/default/files/bbvasoftware/publications/2096\_int21833.pdf (accessed on 17 November 2024).](#A6)

7) [Park, Su-Chan.](#A7) *[Physical Meaning of Principal Component Analysis for Lattice Systems with Translation Invariance](#A7)*[. October 31, 2024. Available online: https://www.researchgate.net/publication/385386437\_Physical\_Meaning\_of\_Principal\_Component\_Analysis\_for\_Lattice\_Systems\_with\_Translational\_Invariance (accessed on 17 November 2024).](#A7)

8) [Sarma, O., Rather, M.A., Shahnaz, S., Barwal, R.S. 2024.](file://C:\\Users\\kkopi\\Downloads\\Sarma, O., Rather, M.A., Shahnaz, S., Barwal, R.S. 2024. Principal Component Analysis of Morphometric Traits in Kashmir Merino Sheep. Journal of Advances in Biology & Biotechnology, volume 27 (9), pp. 362-69. Available from: https:\\www.researchgate.net\\publication\\383479022_Principal_Component_Analysis_of_Morphometric_Traits_in_Kashmir_Merino_Sheep (accessed 18 November 2024).) *[Principal Component Analysis of Morphometric Traits in Kashmir Merino Sheep.](file://C:\\Users\\kkopi\\Downloads\\Sarma, O., Rather, M.A., Shahnaz, S., Barwal, R.S. 2024. Principal Component Analysis of Morphometric Traits in Kashmir Merino Sheep. Journal of Advances in Biology & Biotechnology, volume 27 (9), pp. 362-69. Available from: https:\\www.researchgate.net\\publication\\383479022_Principal_Component_Analysis_of_Morphometric_Traits_in_Kashmir_Merino_Sheep (accessed 18 November 2024).)* [Journal of Advances in Biology & Biotechnology, volume 27 (9), pp. 362-69. Available from: https://www.researchgate.net/publication/383479022\_Principal\_Component\_Analysis\_of\_Morphometric\_Traits\_in\_Kashmir\_Merino\_Sheep (accessed 18 November 2024).](file://C:\\Users\\kkopi\\Downloads\\Sarma, O., Rather, M.A., Shahnaz, S., Barwal, R.S. 2024. Principal Component Analysis of Morphometric Traits in Kashmir Merino Sheep. Journal of Advances in Biology & Biotechnology, volume 27 (9), pp. 362-69. Available from: https:\\www.researchgate.net\\publication\\383479022_Principal_Component_Analysis_of_Morphometric_Traits_in_Kashmir_Merino_Sheep (accessed 18 November 2024).)

9) [Slavkovic, M., Jevtic, D.](#A10) *[Face Recognition Using Eigenface Approach.](#A10)* [SERBIAN JOURNAL OF ELECTRICAL ENGINEERING, Volume 9, No. 1, February 2012, pp. 121-130. Available from https://www.researchgate.net/publication/260880521\_Face\_Recognition\_Using\_Eigenface\_Approach (accessed 18 November 2024).](#A10)

10) [Xu, Z., Guo, F., Ma, H., Liu, X., Gao, L. *On Optimizing Hyperspectral Inversion of Soil Copper Content by Kernel Principal Component Analysis.* Remote Sensing, volume 16, (16). 09 August 2024. Available from: https://www.mdpi.com/2072-4292/16/16/2914 (accessed 18 November 2024](#A9)).

11) [Zhang, H., Srinivasa, R., Yang, X., Ahrentzen, S., Coker, E.S., Alwisy, A,](#A11) *[Factors influencing indoor ai pollution in buildings using PCA – LMBP neural network: A case study of a university campus.](#A11)* [Building and environment, volume 225. November 2022. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0360132322008733](#A11)

[(accessed 18 November 2024).](#A11)

12) [Akinnuwesi, B., Macaulay, B.O., Aribisala, B. *Breast cancer risk assessment and early diagnosis using Principal Component analysis and support vector machine techniques.* Informatics in Medicine Unlocked, Volume 21, 2020. Available from: https://www.sciencedirect.com/science/article/pii/S2352914820306092 (accessed 19 November 2024).](#A12)

13) [Hargreaves, C.A., Mani, C.K. *The Selection of Winning Stocks Using Principal Component Analysis.* American Journal of Marketing Research, volume 1, (3), 2015, pp. 183 – 188. Available from: https://www.researchgate.net/publication/281042621\_The\_Selection\_of\_Winning\_Stocks\_Using\_Principal\_Component\_Analysis (accessed 19 November 2024).](#A13)

14) [Gambardella, C., Parente, R., Ciambrone, A., Casbarra, M. *A Principal Components Analysis – Based Method for the Detection of Cannabis Plants Using Representation data by Remote Sensing.* Knowledge Extractions from data Using Machine Learning, volume 6 (10), 2021. Available from: https://www.mdpi.com/2306-5729/6/10/108 (accessed 19 November 2024).](#A14)

15) [Tsoulos, I.G., Gavrilis, D., Dermatos, E. S*electing and constructing fetures using grammatical evolution.* Pattern Recognition Letters, volume 29 (9), pp. 1358 – 1365. July 2008. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0167865508000664?casa\_token=U8BGSlluRBgAAAAA:KTbdDrgF71wJq\_j-kpQ11HgZWCHrWKDRiTJt7PK2MIH0jDjkNP59dFtC-TXmbcRFFllq0a9wTA (accessed 20 November 2024).](#A15)

16[) Guan, R. *Predicting Forest Fire with Linear Regression and Random Forest.* Highlights in Science, Engineering and Technology, Volume 44 (2023). Available from: https://www.researchgate.net/publication/370220747\_Predicting\_Forest\_Fire\_with\_Linear\_Regression\_and\_Random\_Forest (accessed 20 November 2024).](file://C:\Users\kkopi\Downloads\)%20Guan,%20R.%20Predicting%20Forest%20Fire%20with%20Linear%20Regression%20and%20Random%20Forest.%20Highlights%20in%20Science,%20Engineering%20and%20Technology,%20Volume%2044%20(2023).%20Available%20from:%20https:\www.researchgate.net\publication\370220747_Predicting_Forest_Fire_with_Linear_Regression_and_Random_Forest%20(accessed%2020%20November%202024).)

17) [Nikolov, N., Bothwell, P., Snook,](#A17) *[J. USFS-CSU Joint Venture Agreement Phase 2 (2019-2021): Developing a Gridded Model for Probabilistic Forecasting of Wildland-Fire Ignitions Across the Lower 48 States](#A17)*[. Available from: https://www.fs.usda.gov/rm/pubs\_journals/2022/rmrs\_2022\_nikolov\_n001.pdf](#A17)

[(accessed 20 November 2024).](#A17)

18) [Nikolov, N., Bothwell, P., Snook, J.](#A18) *[Probalistic forecasting of lightning strikes over the Continental USA and Alaska: Model development and verification.](#A18)* [Scientific Journal, Fire (7). Available from: https://research.fs.usda.gov/treesearch/68447](#A18)

[(accessed 20 November 2024).](#A18)