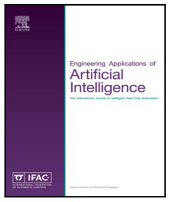
Engineering Applications of Artificial Intelligence 126 (2023) 106772

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Detection of slight variations in combustion conditions with machine learning and computer vision

Pedro Compais [a], Jorge Arroyo [a,*], Miguel Ángel Castán-Lascorz [a], Jorge Barrio [a], Antonia Gil [b]

[a] CIRCE Technology Center, Parque Empresarial Dinamiza Avenida Ranillas, 3D 1ª Planta, 50018, Zaragoza, Spain
[b] Universidad de Zaragoza, IUI mixto CIRCE, Campus Rio Ebro, Mariano Esquillor Gómez, 15, 50018, Zaragoza, Spain

ARTICLE INFO

Keywords:
Combustion monitoring
Slight variation
Machine learning
Computer vision
Color image

ABSTRACT

When monitoring combustion conditions, detecting minor variations, which may be complex even for the human eye, is critical for providing a fast response and correcting deviations. The aim of this study is to detect slight variations in combustion conditions by developing a flame monitoring system using machine learning and computer vision techniques applied to color images. Predictive models are developed for fuel blends with different heating values. The predictive models classify the combustion equivalence ratio based on multiple conditions, using a mean step size of 0.10 between states, a lower value than previously reported in related studies. Three machine learning algorithms are used for each fuel blend: logistic regression, support vector machine, and artificial neural network (multilayer perceptron). These models are fed the statistical, geometrical, and textural features extracted from the color images of the flames. The classification achieves accuracies from 0.78 to 0.97 in the detection of slight variations in the combustion conditions for all heating values. Thus, the monitoring system developed in this study is a promising alternative for implementation on an industrial scale and quick detection of changes in combustion conditions.

1. Introduction

Combustion disturbances may shift controlled combustion conditions towards abnormal operation regimes, leading to flashback or extinction of the flame in the worst-case scenario. Measuring minor deviations in combustion conditions from standard combustion operations can enable an early detection of abnormal operation regimes. This early detection is necessary to quickly adjust the process, thereby reducing the operating time under suboptimal conditions and other efficiency problems. Machine learning (ML) techniques can be used to enable advanced combustion monitoring.

Currently, ML techniques are employed in many fields, such as agriculture (Lawal, 2021), surveillance (Matkvoic et al., 2022), biochemical engineering (Mowbray et al., 2021; Roy, 2022), heat pipes (Wang et al., 2021), power systems (Vaish et al., 2021), control systems (Singer and Cohen, 2021), and combustion engines (Aliramezani et al., 2022). ML techniques are employed to analyze data for obtaining insights and achieving higher levels of automation in data analysis. Thus, ML provides powerful tools for analyzing large datasets and addressing problems that are extremely complex or unviable with traditional data analytics. Within the field of combustion, ML has been recently used to optimize the trade-off between emissions and efficiency (Cheng et al., 2018) and to predict multiple characteristics, such as dynamic and steady behavior (Jung et al., 2023), emissions (González-Espinosa et al., 2020; Park et al., 2022), fuel oil viscosity (Ibargüengoytia et al., 2013), and long-term furnace temperature (Quesada et al., 2021). Flame images have been used to predict the operating conditions related to the air–fuel equivalence ratio (ER), such as the air ratio (Bai et al., 2017), O_2 concentration (Yang et al., 2022), and combustion regimes (Abdurakipov et al., 2018; Han et al., 2020, 2021). Combustion conditions were examined in discrete ranges at different values (steps). Mean step widths (step size) of 0.20 (Han et al., 2020) and 0.35 (Abdurakipov et al., 2018) were used to evaluate the influence of the ER on combustion performance. However, the ER can be measured with a higher level of detail by increasing the number of steps (i.e., reducing the step size).

Several ML algorithms have been used to predict combustion conditions, such as logistic regression (LR) (Abdurakipov et al., 2018; Han et al., 2021; Hanuschkin et al., 2021), Gaussian processes (Han et al., 2021), decision trees (Han et al., 2021; Hanuschkin et al., 2021), *k*-nearest neighbor (Bai et al., 2017; Abdurakipov et al., 2018),

Abbreviations: ANN, artificial neural network; ANOVA, analysis of variance; BFG, blast furnace gas; CV, cross-validation; DL, deep learning; ER, equivalence ratio; GLCM, grey level co-occurrence matrix; IF, image feature; LHV, low heating value; LR, logistic regression; ML, machine learning; MLP, multilayer perceptron; PCA, principal component analysis; SVM, support vector machine

* Corresponding author.
E-mail address: jarroyo@fcirce.es (J. Arroyo).

https://doi.org/10.1016/j.engappai.2023.106772
Received 21 December 2022; Received in revised form 23 May 2023; Accepted 4 July 2023
Available online xxxx

linear discriminant analysis (González-Espinosa et al., 2020), support vector machines (SVMs) (Bai et al., 2017; Abdurakipov et al., 2018; Han et al., 2020, 2021), and artificial neural networks (ANNs) (Bai et al., 2017; Abdurakipov et al., 2018; González-Espinosa et al., 2020; Han et al., 2020, 2021; Hanuschkin et al., 2021; Yang et al., 2022). Predictive models are typically fed with relevant features obtained from flame images, which correlate with the combustion variables. Many of these features are extracted from flame images by using computer vision techniques. Features are typically based on statistical (González-Cencerrado et al., 2012, 2013; Sun et al., 2013; González-Cencerrado et al., 2015; Sun et al., 2015; Mathew et al., 2016; Bai et al., 2017; Katzer et al., 2017; González-Espinosa et al., 2020; Compais et al., 2022a,b; Zhu et al., 2023), geometrical (Sun et al., 2013, 2015; Katzer et al., 2017; Hanuschkin et al., 2021; Liu et al., 2021; Compais et al., 2022a), and textural measures (Bai et al., 2017; Compais et al., 2022a; Yang et al., 2022).

Owing to the many available features, most applications choose a limited data subset tailored to the case study. To compress and visualize feature information, some researchers have employed principal component analysis (PCA) (Bai et al., 2017; Abdurakipov et al., 2018; Hanuschkin et al., 2021; Yang et al., 2022). However, PCA does not consider the relevance of features for predicting a target. Thus, other techniques are preferred because PCA is not recommended for addressing overfitting (Hanuschkin et al., 2021; Yang et al., 2022). After the predictive models are developed, their performance is usually evaluated using accuracy, F1-score or R^2 metrics, and training-test split and cross-validation (CV) methods. However, predictive models may suffer from overfitting and provide overly optimistic performance results. Several alternatives, such as feature selection, regularization term, and CV, can be used to address overfitting and obtain more robust measures (Bai et al., 2017; Abdurakipov et al., 2018; Han et al., 2020, 2021; Hanuschkin et al., 2021; Quesada et al., 2021). Features can be manually or automatically selected by using other techniques, such as Pearson's correlation coefficient or analysis of variance (ANOVA). CV can be employed for model evaluation and hyperparameter tuning. However, using a unique CV for both tasks may result in overfitting (Cawley and Talbot, 2010). Two different CVs can address this risk: one for hyperparameter tuning and the other for performance evaluation (Cawley and Talbot, 2010; Hanuschkin et al., 2021). Thus, the CV for hyperparameter tuning (inner CV) is nested under the CV for performance evaluation (outer CV), resulting in a nested CV. Owing to model overfitting, using a non-nested CV instead of a nested CV may provide overly optimistic results for model performance. For example, a 13% accuracy reduction was reported in classification models using nested CV instead of non-nested CV (Abdulaal et al., 2018). Thus, the validation procedures used should be considered when comparing quantitative results from other studies. However, nested and non-nested CVs generally result in the selection of the same model for prediction applications (Wainer and Cawley, 2018).

This study presents an advanced monitoring system based on ML and computer vision to detect minor variations in the combustion conditions. Several ML models are developed and tested on a laboratory-scale burner using different proportions of pure methane (CH_4), a baseline fuel, and blast furnace gas (BFG), a lean fuel. BFG, which is a by-product of the integrated steelmaking sector and decreases the low heating value (LHV) of the gas mixture, thereby leading to different burner behaviors under premixed combustion conditions. Using BFG as fuel in the steel sector is encouraged to improve energy efficiency and reduce fossil fuel consumption and global CO_2 emissions (Cuervo-Piñera et al., 2018). Thus, advanced combustion monitoring systems based on image sensors are helpful tools for this purpose.

This work presents the development of predictive models for ER classification with a significantly high level of detail, using a mean step size of 0.10 between consecutive combustion conditions. This accuracy exceeds the human eye's sensibility and implies accurate control of combustion processes (Bai et al., 2017). The methodology includes three characteristics that had not been implemented together before and whose separate use is scarcely reported in the field of combustion monitoring. First, whereas most combustion studies only include one or two image features (IFs), in this work, statistical, geometrical, and textural IF are extracted to achieve a complete combustion characterization. Second, ANOVA F-tests are performed to automatically select IFs for training predictive models. This approach is implemented to overcome the potential issue of overfitting, which is a significant obstacle to minor ER changes.

Finally, while other studies involved the use of less robust methods, such as a unique CV for model evaluation and hyperparameter turning or no CV at all, we use nested CV for model evaluation and hyperparameter tuning. Here, the hyperparameters for predictive models are automatically defined with the nested CV. The model accuracy is measured for each hyperparameter combination in the inner CV, and the combination with the highest accuracy is selected for the outer CV. The objective of implementing a nested CV is the same as that of the ANOVA *F*-tests: the reduction of overfitting in detecting slight ER variations. Predictive models are developed using three different ML algorithms: LR, SVM, and ANNs with multilayer perceptron (MLP). The performance of the predictive models is evaluated to study their behavior and compare the differences between the models, ML algorithms, ER classes, and fuel blends.

2. Material and methods

2.1. Experimental setup

Experimental tests were performed in a combustion chamber equipped with a premixed gas fuel burner with a maximum power of 20 kW_{th}. Fig. 1 shows a schematic of the experimental setup. The burner comprised two separate fuel and air inlets with diameters of 25 mm and 10 mm. Air and fuel were premixed in a plenum inside the burner, and the mixture left through a 100-mm-diameter header and a pattern of 5-mm-diameter holes.

The burner was fed from bottles of gaseous fuel through two independent lines. Each line was designed to feed gaseous fuel with different heating values. The fuel mixtures were prepared by using a gas supplier based on the composition and quality requirements. A compressor supplied the combustion air, and an ITV2000 electropneumatic regulator (SMC España S.A., Spain) controlled the pressure (and thus the airflow rate). The airflow rate was measured before burner connection using an SD6000 airflow switch (IFM Electronic GmbH, Germany). The pressure control and airflow rate were digitized using a data acquisition system and computer. Flame color images were acquired by using a DFK 33GX174 color camera (The Imaging Source Europe GmbH, Germany) with a IMX174LQJ sensor (Sony Europe, Netherlands) of 2.3 MP.

Three fuel blends with different LHV values (Table 1) were tested during the experiment. Fuels were selected based on interest in using BFG, a low-calorific power gas, in the steel sector. Blends with a high percentage of BFG led to increased combustion instability owing to the composition of the inert gases. Pure CH_4 was employed as the baseline for combustion with the highest LHV (MIX1). In contrast, the pure BFG exhibited the lowest LHV (MIX3). Finally, a fuel blend composed of 30% vol. CH_4 and 70% vol. BFG was studied as an intermediate LHV scenario (MIX2). The latter is relevant in the steel industry for the valorization of BFG in reheating furnaces (Caillat, 2017; Cuervo-Piñera et al., 2017). For each fuel blend, the combustion conditions were modified by changing the airflow rate and operating with different ERs at a fixed power of 5.5 kW_{th}. Further details regarding the experimental procedure can be found in Compais et al. (2022a).

2.2. Methods

Here, predictive models focused on the innovative detection of minor ER variations, providing greater detail in estimating the ER than

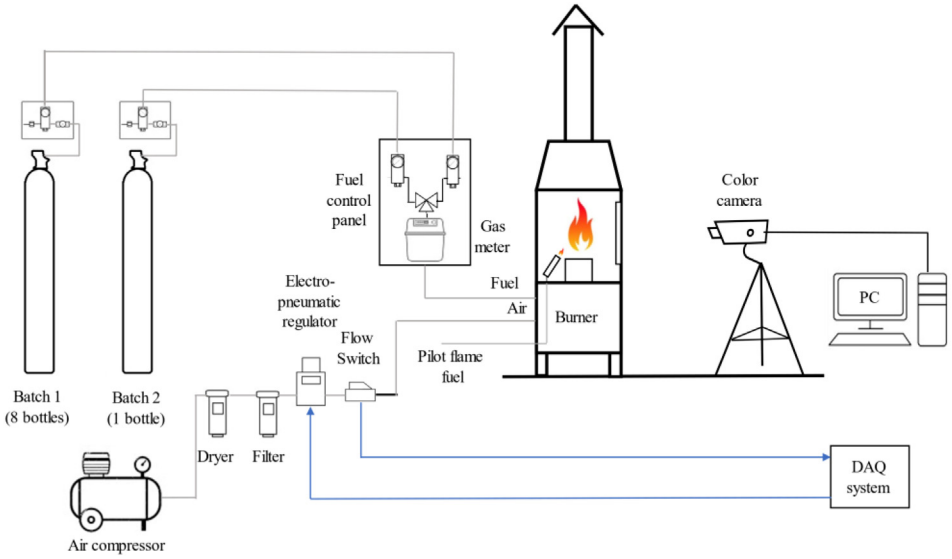


Fig. 1. Scheme of the combustion laboratory. DAQ: data acquisition.

Table 1
Composition of fuel blends (Compais et al., 2022a).

Fuel blend	MIX1	MIX2	MIX3
[CH_4] (%vol.)	100	28	–
[H_2] (%vol.)	–	3	4
[CO] (%vol.)	–	16	22
[CO_2] (%vol.)	–	16	22
[N_2] (%vol.)	–	37	52
LHV (MJ/kg)	50.0	10.8	2.8

Table 2
Summary of the experimental tests.

Test	MIX1	MIX2	MIX3
ER_1	1.41	1.13	0.91
ER_2	1.43	1.27	0.94
ER_3	1.51	1.37	1.09
ER_4	1.64	1.41	1.11
ER_5	1.76	1.57	1.24
ER_6	1.88	1.67	–
ER_7	2.01	1.78	–
ER_8	–	1.91	–

in previous studies. Combustion measurements from the experiment conducted by Compais et al. (2022a) were used as input to develop the predictive models. The ER was computed as the ratio between the actual and stoichiometric air–fuel ratios. The mean ER step sizes were 0.10. The ER range was defined for each fuel blend based on its specific flame stability. For example, an ER of 2.01 was achieved for MIX1 but not for MIX2 or MIX3, owing to the extinction of the flame. Table 2 presents the ERs tested for each fuel blend. ERs range from 1.41 to 2.01 (MIX1), 1.13 to 1.91 (MIX2), and 0.91 to 1.24 (MIX3).

The camera acquired the flame images for 6 min after reaching steady conditions for each test (ER class). The camera achieved a frame rate of 12 fps, capturing 4320 images per test. Figs. 2, 3 and 4 show sample images of the combustion regimes for MIX1, MIX2, and MIX3. The ER variations in the flames were so small that they were not visible to the naked human eye. This effect was attributed to using a mean step size of 0.10 ER. Although slight image variations might seem irrelevant at a lab scale, they might result in high volumes of natural gas that were not burned on an industrial scale. In this respect, fine detection of these slight variations could provide energy savings to the steel industry.

Flame images captured during the tests were processed to extract their features. Although all the main feature types for flame images

Table 3
Statistical IF extracted from flame pixels of each color channel.

Feature magnitude	Equation
Mean (μ)	$\frac{1}{P}\sum_{p=1}^{P} x(p)$
Standard deviation (σ)	$\sqrt{\frac{1}{P}\sum_{p=1}^{P}(x(p)-\mu)^2}$
Skewness (s)	$\frac{\frac{1}{P}\sum_{p=1}^{P}(x(p)-\mu)^3}{\sigma^3}$
Kurtosis (k)	$\frac{\frac{1}{P}\sum_{p=1}^{P}(x(p)-\mu)^4}{\sigma^4}$

(statistical, geometrical, and textural) were not included in most combustion studies, they were integrated for comparison in this study. Four statistical features (mean, standard deviation, skewness, and kurtosis), 13 textural features (selected from Haralick et al., 1973), and five geometrical features (area, centroid horizontal coordinate, centroid vertical coordinate, width, and height) were computed. Processing was applied to each color channel to obtain 22 IF per color channel, for 66 IF per color image. Tables 3, 4, and 5 list the selected IFs with their formulations. The equations of the statistical features were referred to as monochrome images of P pixels, where $x(p)$ was the value of pixel p. In textural features, $p(i,j)$ referred to elements in row i and column j of a normalized GLCM. The GLCM had N rows and N columns, where N was defined as the number of gray values in the monochrome image. Geometrical features were computed for binary images of P pixels (with R rows and C columns), with $b(c,r)$ as the binary value (zero or one) of pixel p located in column c and row r. Moreover, C_s and R_s were the horizontal and vertical coordinates of the pixels with binary values of one, respectively.

Before the IF was extracted, the flame images were preprocessed. First, to remove sensor electrical noise, the dark camera signal was subtracted from the flame images (González-Cencerrado et al., 2012, 2013; Huang et al., 2015). The flame pixels were then segmented using thresholding (Mathew et al., 2016; Katzer et al., 2017). The threshold was automatically selected by applying Otsu's method to maximize the variance between the two-pixel classes (Otsu, 1979). The statistical and textural features were extracted. Using Otsu's thresholding, a reduced number of image pixels that were distant from the main flame body were erroneously classified as flame pixels. Although they barely affected the statistical and textural characteristics, these pixels significantly influenced the calculation method for the geometrical features. Therefore, the morphological transformation of erosion (Sreedhar and

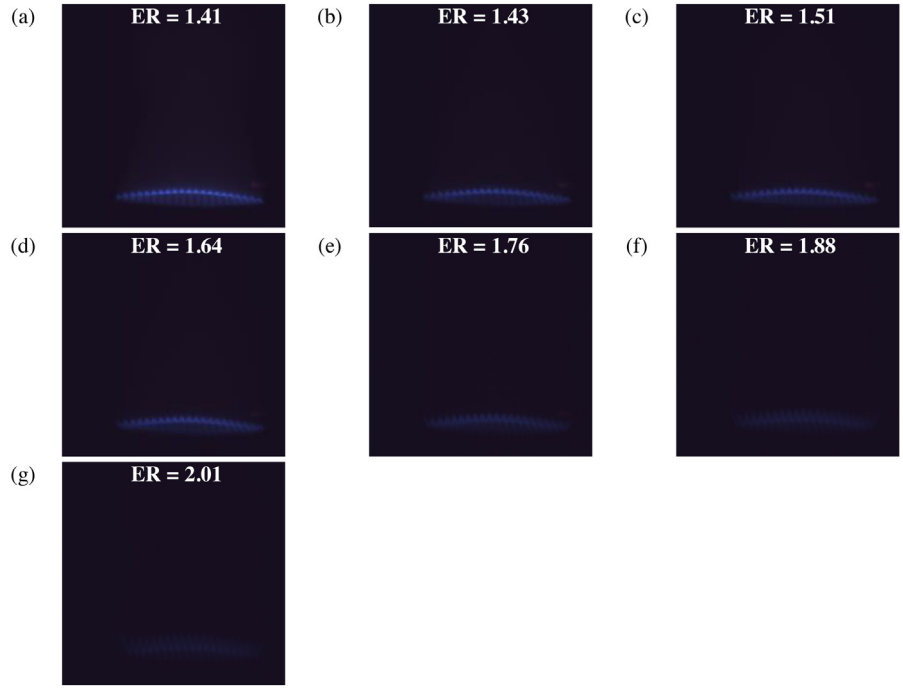


Fig. 2. Sample images of flames for MIX1 and ERs of (a) 1.41, (b) 1.43, (c) 1.51, (d) 1.64, (e) 1.76, (f) 1.88, and (g) 2.01.

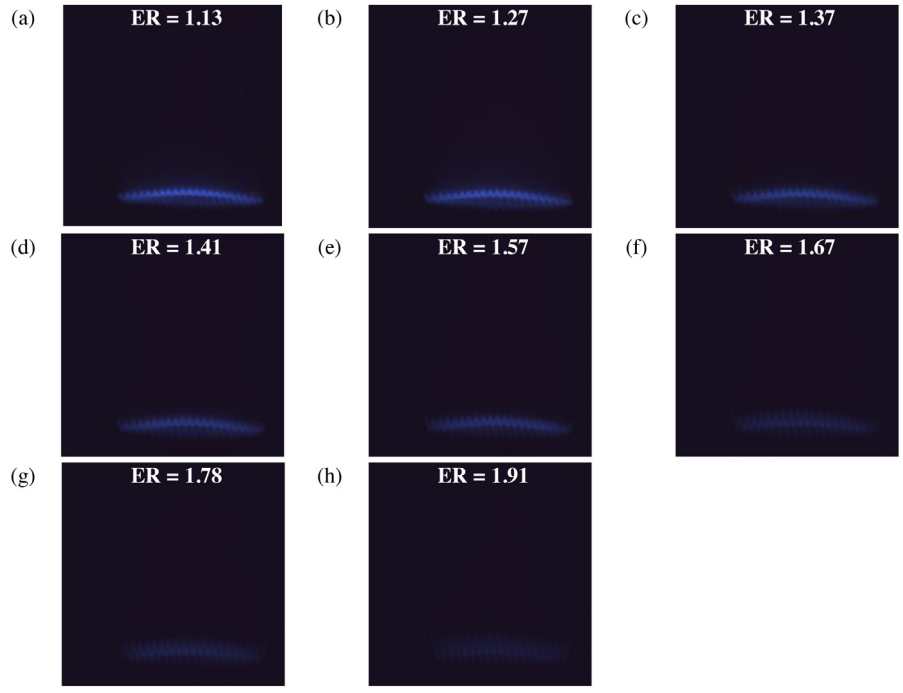


Fig. 3. Sample images of flames for MIX2 and ERs of (a) 1.13, (b) 1.27, (c) 1.37, (d) 1.41, (e) 1.57, (f) 1.67, (g) 1.78, and (h) 1.91.

Panlal, 2012) was used to discard them. Erosion was applied using a 3 × 3-pixel kernel. Finally, geometrical features were extracted.

Based on the same methodology, different predictive models were developed to detect slight ER variations in three fuel blends (MIX1, MIX2, and MIX3). Each image was labeled with its corresponding fuel blend and ER. Flame color images, fuel blends, and ER labels were used as the datasets. Each fuel blend was tested for a discrete group of ERs and predictive models were developed to classify each ER label. These predictive models estimated the ERs related to the images based on the extracted IF. The behavior of several ML algorithms was analyzed for the ER classification of fuel blends. Predictive models were developed using ML algorithms with different characteristics, namely LR, SVM, and ANN. An MLP with a unique hidden layer of 100 neurons was used for the latter. The ML methodology in this study for each fuel blend is shown in Fig. 5.

The overall ML process for the dataset of a fuel blend is summarized as follows, and specific steps of the method are described in more detail. The dataset of each fuel blend was randomly shuffled and split into training and test sets with 70% and 30% of the samples, respectively. These sets were stratified to include a similar proportion of the ER classes. Each IF's mean and standard deviation in the training set were computed to standardize the dataset. Based on those values,

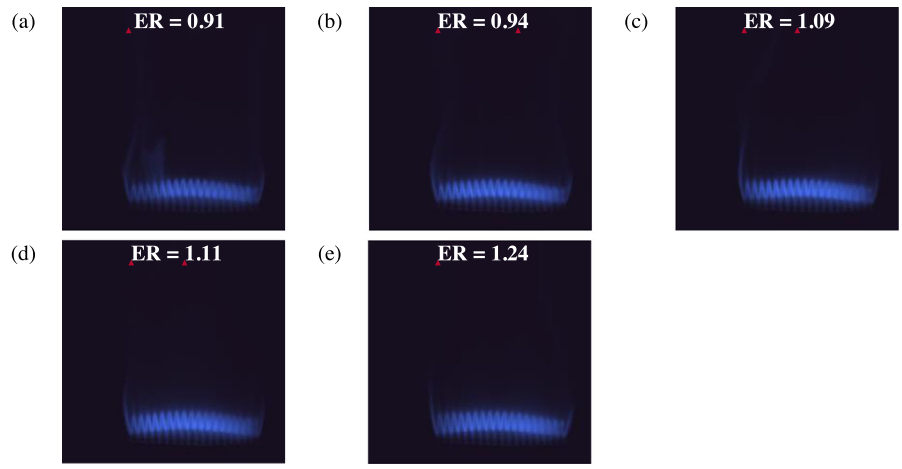


Fig. 4. Sample images of flames for MIX3 and ERs of (a) 0.91, (b) 0.94, (c) 1.09, (d) 1.11, and (e) 1.24.

Table 4
Textural IF extracted from flame pixels of each color channel.

Feature magnitude	Equation
Angular second moment (f_1, energy)	$\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)^2$
Contrast (f_2)	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i-j)^2 p(i,j)$
Correlation (f_3)	$\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
Sum of squares (f_4, variance)	$\sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu_x)^2 p_{ij}$
Inverse difference moment (f_5)	$\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{p_{ij}}{1+\lvert i-j\rvert}$
Sum average (f_6)	$\sum_{k=2}^{2N} k p_{x+y}(k)$
Sum variance (f_7)	$\sum_{k=2}^{2N} (k-\mu_{x+y})^2 p_{x+y}(k)$
Sum entropy (f_8)	$-\sum_{k=2}^{2N} p_{x+y}(k)\log p_{x+y}(k)$
Entropy (f_9)	$-\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)\log p(i,j)$
Difference variance (f_{10})	$\sum_{k=0}^{N-1} (k-\mu_{x-y})^2 p_{x-y}(k)$
Difference entropy (f_{11})	$-\sum_{k=0}^{N-1} p_{x-y}(k)\log p_{x-y}(k)$
Information measure of correlation I (f_{12}, IMC1)	$\frac{HXY-HXY1}{max(HX,HY)}$
Information measure of correlation II (f_{13}, IMC2)	$\sqrt{1-\exp[-2(HXY2-HXY)]}$

Table 5
Geometrical IF extracted from flame pixels of each color channel.

Feature magnitude	Equation
Area (a)	$\sum_{r=1}^{R}\sum_{c=1}^{C} b(c,r)$
Centroid horizontal coordinate (c_x)	$\frac{\sum_{r=1}^{R}\sum_{c=1}^{C} cb(c,r)}{a}$
Centroid vertical coordinate (c_y)	$\frac{\sum_{r=1}^{R}\sum_{c=1}^{C} rb(c,r)}{a}$
Width (w)	$\max(C_s)-\min(C_s)$
Height (h)	$\max(R_s)-\min(R_s)$

training and test sets were standardized. To tackle the overfitting of the predictive models, the training set was analyzed by ANOVA F-tests. The ten IFs with the best variance results were selected and employed as input for predictive models for both training and test sets. Three nested CVs (one per ML algorithm) tuned the hyperparameters and evaluated the performance of each ML algorithm, selecting the best alternative in the end. The chosen ML algorithm was trained with the selected hyperparameters and the training set, and its performance was analyzed, computing its accuracy for the test set. Accuracy was defined as the ratio of the number of correct predictions to the total number of predictions. PCA was not used in the ML method because training time and storage limitations were not critical in this study.

Regarding the ANOVA F-tests, the selected IFs were the ten variables with the highest variance for the ER classes of each fuel blend in the training set. F-tests use a Fisher–Snedecor distribution, and in the case of the ANOVA F-tests, the hypothesis was the dependence of an image feature on the ER class for a specific fuel blend. A confidence level of 0.05 was selected to evaluate this hypothesis, and the F-values, critical F-values, and p-values of the 66 IFs were computed and compared. In the ANOVA F-test of an IF, the hypothesis was supported if its F-value was higher than the critical F-value for the specific fuel blend and the p-value was lower than the confidence level.

Nested CVs were applied with stratification to the training set, with an outer and inner CV of ten and five folds, respectively. The outer CV split the training set into ten training- and validation-subset pairs. Fig. 6 summarizes the ML method for a split i of an outer CV.

For a training subset, an inner CV was applied to define the hyperparameters of the predictive model. Next, the model was trained with the training subset, and its performance was evaluated by calculating its training and validation accuracies, learning curve, and validation confusion matrix. Training and validation accuracies were measured for different subset sizes to compute the learning curves. These sizes were defined as 1%, 25%, 50%, 75%, and 100% of the samples in the outer CV fold. After the ten splits were evaluated, their metrics were averaged, and the mean accuracy was used to select the ML algorithm. The inner CV split each training subset into another five pairs of training and validation subsets. For each pair, every combination of hyperparameters was evaluated. The procedure for the split j and the combination of hyperparameters k is shown in Fig. 7.

For an inner CV split j and a combination of hyperparameters k, the predictive model was trained with the training subset j, and its accuracy was computed for the validation subset j. Values for the five splits were averaged to define the combination of hyperparameters with the highest accuracy. Three values of the regularization term (0.1, 1, and 10) were tested for the three ML algorithms to address the overfitting problem. Moreover, for the SVM, three different kernels were tested (linear, polynomial, and radial basis functions).

The specific code to extract the IF and develop predictive models was developed using Python as the programming language (version 3.7). Several libraries were used for the code: OpenCV, Scikit-learn, NumPy, SciPy, Mahotas, and Pandas.

3. Results and discussion

Several comparative analyses were performed to evaluate the detection of slight ER variations using predictive models. First, the variance of IF with the ER classes was analyzed using the ANOVA F-tests. This assessment compared the subsets of the IF automatically selected for each fuel blend. Apart from the specific IF that formed each subset, the relevance of each subset's color channels, and feature types was checked. Next, the validation accuracy achieved by the predictive models and the effects of the ML algorithm were tested against other related

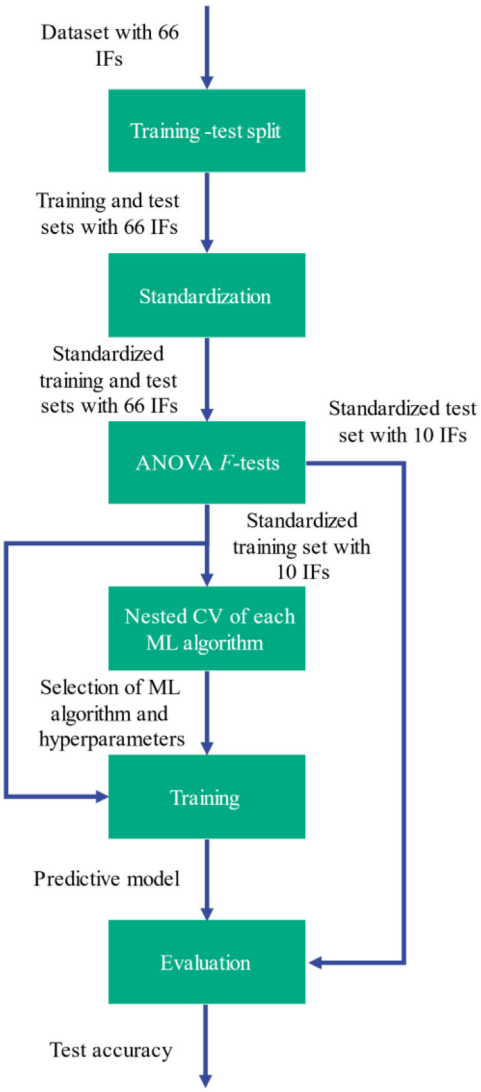


Fig. 5. ML method for each fuel blend.

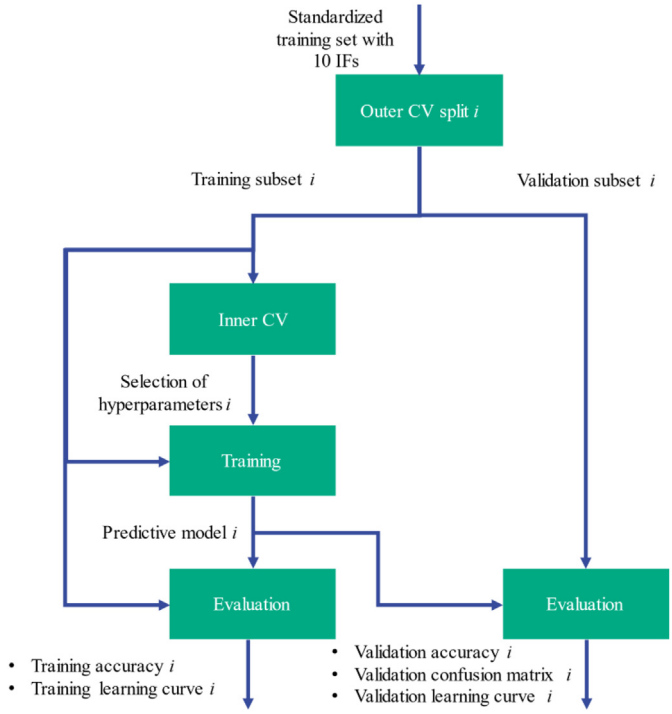


Fig. 6. ML method for the split i of an outer CV.

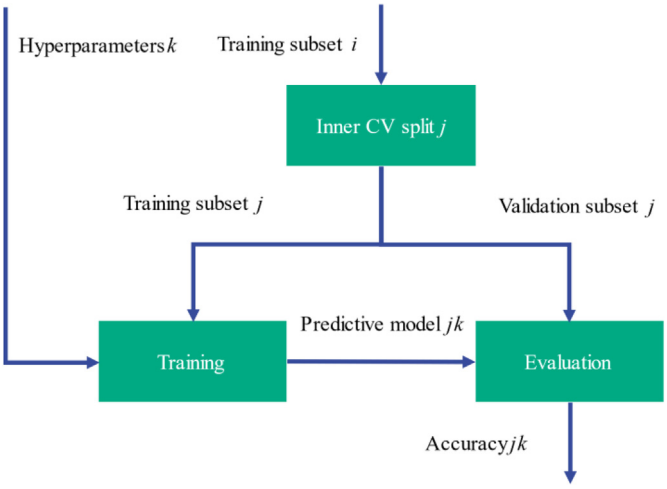


Fig. 7. ML method for the split j of an inner CV and the combination of hyperparameters k.

studies. Also, confusion matrices were computed to evaluate the effect of ER class on the prediction models. Finally, the fit of the predictive models was analyzed using learning curves and by comparing the training, validation, and test accuracies.

3.1. Variance of the IF with the ER classes

The first analysis measured the variance of IF with the ER classes using the ANOVA *F*-test. The critical *F*-value was computed for each fuel blend, with a confidence level of 0.05. Fuel blends had critical *F*-values of 2.10 (MIX1), 2.01 (MIX2), and 2.37 (MIX3). All IF had *F*-values that were at least two orders of magnitude higher than the critical *F*-value of the fuel blend. Also, all *p*-values were lower than the confidence level. Thus, the mean values of each IF were affected by the ER class regardless of the fuel heating value. IF were ordered based on their *F*-values, and the subset of the 10 IF with the highest *F*-values was selected to develop the predictive models. The chosen IF are indicated with checkmarks (Tables 6, 7, and 8).

Three subsets of 10 IF were selected, with one subset for each fuel blend. Some visual characteristics were repeated between subsets. In particular, the IF group formed by the three subsets included only 24 different IF from a total computed of 66. The subsets for MIX1 and MIX2 shared four IF: the standard deviation of the green and blue channels, difference entropy of the green channel, and centroid vertical coordinate of the blue channel, which was a feature shared for all fuel blends. The number of IF was computed for each color channel (Fig. 8[a]) and feature types (Fig. 8[b]). The subsets of IF depended on the fuel blend.

For MIX1, only the IF from the green and blue channels were selected. However, in the cases of MIX2 and MIX3, the IFs from the three-color channels were included. Moreover, three feature types were included in selecting IF for MIX1 and MIX2. Nevertheless, only the textural and geometrical features were chosen for MIX3.

3.2. Prediction of slight variations in the ER

After analyzing the variance of the IF with the ER, for each fuel blend, a subset of ten features was selected, which were used for the

Table 6
IF selected from the red color channel for each fuel blend.

Feature type	Feature magnitude	MIX1	MIX2	MIX3
Statistical	Standard deviation (σ)		✓	
Textural	Contrast (f_2)		✓	
Geometrical	Area (a)			✓
Geometrical	Centroid vertical coordinate (c_y)			✓

Table 7
IF selected from the green color channel for each fuel blend.

Feature type	Feature magnitude	MIX1	MIX2	MIX3
Statistical	Standard deviation (σ)	✓	✓	
Textural	Contrast (f_2)		✓	
Textural	Correlation (f_3)			✓
Textural	Sum of squares (f_4, variance)		✓	
Textural	Inverse difference moment (f_5)	✓		
Textural	Difference entropy (f_{11})	✓	✓	
Textural	Information Measure of Correlation II (f_{13}, IMC2)			✓
Geometrical	Area (a)			✓
Geometrical	Centroid vertical coordinate (c_y)	✓		✓

Table 8
IF selected from the blue color channel for each fuel blend.

Feature type	Feature magnitude	MIX1	MIX2	MIX3
Statistical	Standard deviation (σ)	✓	✓	
Statistical	Skewness (s)	✓		
Textural	Contrast (f_2)		✓	
Textural	Correlation (f_3)			✓
Textural	Sum of squares (f_4, variance)		✓	
Textural	Inverse difference moment (f_5)	✓		
Textural	Difference entropy (f_{11})	✓		
Textural	Information Measure of Correlation I (f_{12}, IMC1)	✓		
Textural	Information Measure of Correlation II (f_{13}, IMC2)			✓
Geometrical	Area (a)			✓
Geometrical	Centroid vertical coordinate (c_y)	✓	✓	✓

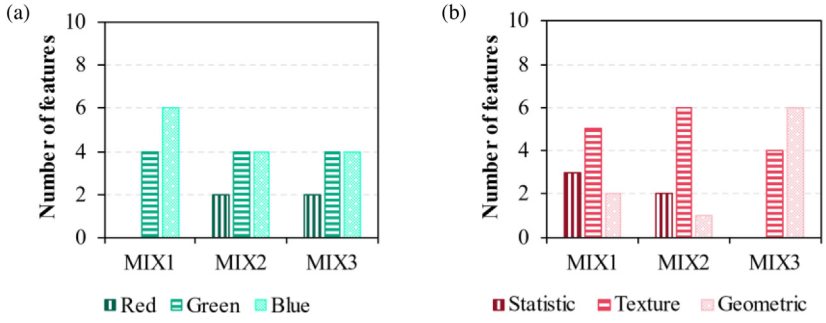


Fig. 8. Number of IF selected for fuel blend, based on (a) their color channel and (b) their feature type.

input and development of predictive models. Three predictive models were developed for each fuel blend using different ML algorithms (LR, SVM, and MLP), and were trained and validated using nested CV. The hyperparameters of the ML algorithms were tuned in this nested CV. The validation accuracies of the predictive models were averaged over the nested CV, and Fig. 9 shows the results.

The predictive models showed similar results for MIX1 and MIX2 expression. ER was estimated using classes with a mean step size of 0.10 (MIX1) and 0.11 (MIX2), achieving validation accuracies between 0.95 and 0.97. In contrast, the ER steps for MIX3 had a lower mean step size (0.08), and validation accuracies of approximately 0.78 were achieved. In summary, the validation accuracies ranged between 0.78 and 0.97, which are typical values for classification models related to ER conditions. For example, Bai et al. (2017) measured accuracies between 0.75 and 0.93 to predict air ratios. However, Han et al. (2021) achieved accuracies of between 0.96 and 1.00 and 0.65 and 0.99 (Han et al., 2020) for the classification of combustion states. Also, Abdurakipov et al. (2018) reported accuracies of 0.89 and 0.98.

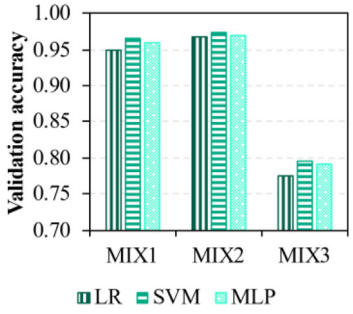


Fig. 9. Validation accuracy of the predictive models developed with different algorithms.

The results were compared in more detail with those of works on similar classification tasks based on combustion conditions with

different ER. Han et al. (2020) and Abdurakipov et al. (2018) used mean step sizes of 0.20 and 0.35, respectively, whereas the present work predicted ER conditions using a mean step size of 0.10, which is at least two times lower. The use of larger mean step sizes facilitates the image classification task because image flames present significant differences that could be perceived by the human eye, as shown when analyzing figures in their works. However, the classification task in this work is more difficult because of the smaller mean step size, which provided image differences hardly perceived by the human eye. To address this challenge, we proposed a novel methodology for previous combustion studies. In summary, using smaller ER steps and a nested CV, the current study predicted ER conditions with a considerably high level of detail, with accuracies like previous studies.

3.3. Effect of the ML algorithm

Results (Fig. 9) show that using a different ML algorithm barely affected the validation accuracy of the predictive models regardless of the fuel blend. LR, SVM, and MLP models achieved validation accuracies with a maximum deviation of 3%. SVM provided the highest accuracies for the three fuel blends (0.967, 0.973, and 0.795). Previous studies measured variations in accuracy below 2% between SVM and ANN (Bai et al., 2017) and LR and SVM (Abdurakipov et al., 2018). In the work of Han et al. (2020), the highest accuracy was achieved by SVM compared with ANN. Nevertheless, high deviations (< 19%) between LR, SVM, and ANN have been reported in some cases (Bai et al., 2017; Abdurakipov et al., 2018; Han et al., 2020). Moreover, ANN achieved better results than SVM in the studies by Bai et al. (2017) and Abdurakipov et al. (2018). Therefore, the behavior of the ML algorithms may be dependent on the case conditions, and the current research achieved similar results to those of previous studies.

3.4. Effect of the ER class

As Fig. 9 shows, the validation accuracy decreased significantly for MIX3. To analyze this behavior in detail, the confusion matrixes of the predictive models were analyzed. Results are shown (Fig. 10) for the SVM model, which provided the highest validation accuracy.

For MIX1 and MIX2, confusion matrixes showed similar results. Regardless of the ER class, the predictive models correctly estimated the ER value for most samples (95% at minimum). Incorrect classifications occurred only for a few samples with consecutive ER classes, such as 1.43 and 1.51 in the case of MIX1 (Fig. 10[a]). However, the behavior of the predictive model was different for MIX3, where the ER estimation exhibited lower accuracies when distinguishing between class pairs of 0.91–0.94 and 1.09–1.11 (Fig. 10[c]). These conditions had the smallest ER variations (0.03 and 0.02), together with 1.41–1.43 for MIX1. However, the latter case was more accurate despite its low ER variation. Due to the lower combustion stability caused by its lower heating value, the classification could be more complex for MIX3 than for MIX1. Also, it is observed that when the ER increased between 0.91 and 1.24, the validation accuracy for MIX3 was higher. These variations could be related to combustion stability changes, as reported by Zheng et al. (2021).

3.5. Fit of the predictive models

To evaluate the predictive models in more detail, learning curves were computed. Fig. 11 shows the learning curves for the predictive models developed using the SVM, which provided the highest accuracies.

For each fuel blend, the validation accuracy increased with increasing subset size, approaching the same value as the training accuracy. This behavior is considered a good fit for the predictive models, which did not suffer from underfitting or overfitting. Furthermore, as the predictive models were stable at a subset size of 25%, similar results

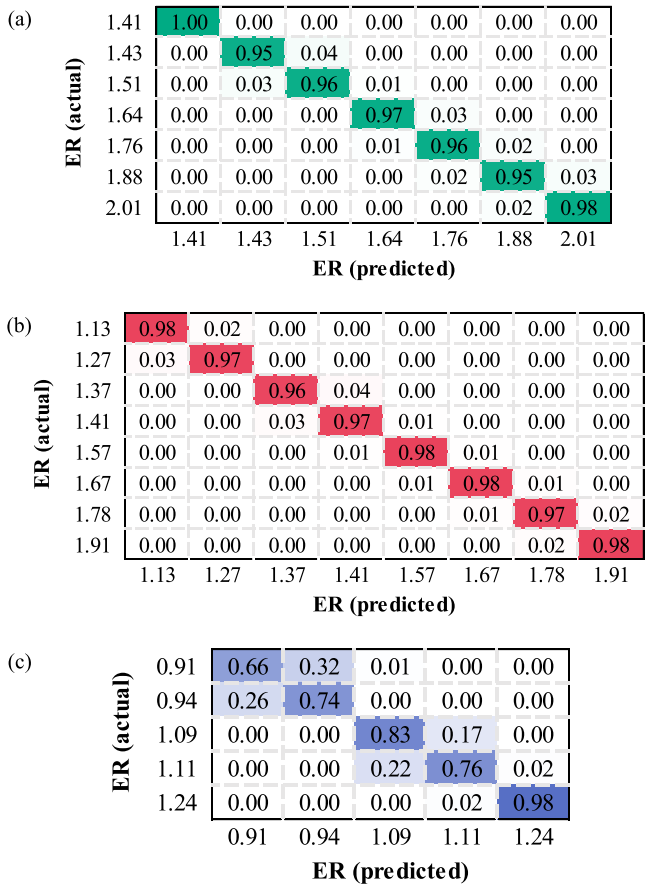


Fig. 10. Confusion matrixes of the predictive models developed with the SVM algorithm for (a) MIX1, (b) MIX2, and (c) MIX3.

could be achieved by acquiring only 1080 images per class, reducing the test duration from 360 s (Compais et al., 2022a) to 90 s, or the frame rate from 12 (Compais et al., 2022a) to 3 fps.

To provide a final test for the ER classification, predictive models with the best results were trained with the entire training set and evaluated with the test set, which had not been used to develop the predictive models. Only the SVM models were evaluated in this step, as they achieved the highest accuracy. The hyperparameters of the SVM models were defined based on the best results obtained in the hyperparameter tuning: a regularization term of 10 for the three fuel blends, radial basis function kernel for MIX1 and MIX3, and linear kernel for MIX 2. Their accuracies were computed to evaluate the SVM models using the test set. Fig. 12 shows the accuracies for the SVM model, including the previous training and validation accuracies and the new test accuracies.

The accuracies achieved using the test set were like those of the training and validation sets. Therefore, the SVM models exhibited acceptable behavior for previously unseen flame images.

4. Conclusions

This paper presented a novel methodology for detecting slight variations in combustion conditions. The fuel blends used here are pure CH_4, 30% vol. CH_4/70% vol. BFG, and pure BFG. The combustion was analyzed using a laboratory-premixed burner at a fixed thermal power of 5.5 kW. The developed methodology was based on extracting statistical, geometrical, and textural features from flame images, their automatic selection with ANOVA *F*-tests, the automatic selection of hyperparameters, and the robust performance evaluation for predictive

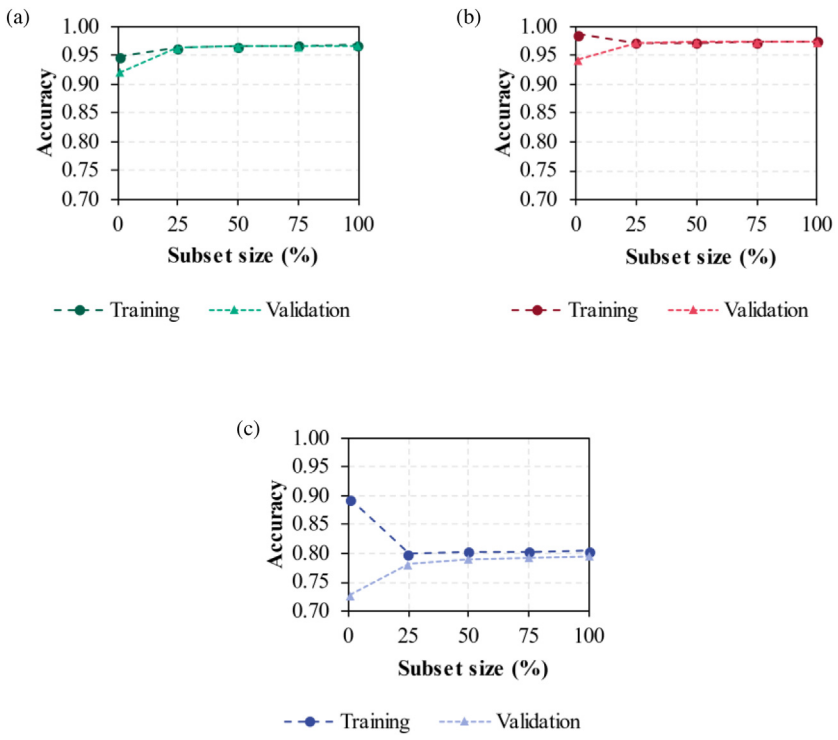


Fig. 11. Learning curves of the SVM model for (a) MIX1, (b) MIX2, and (c) MIX3.

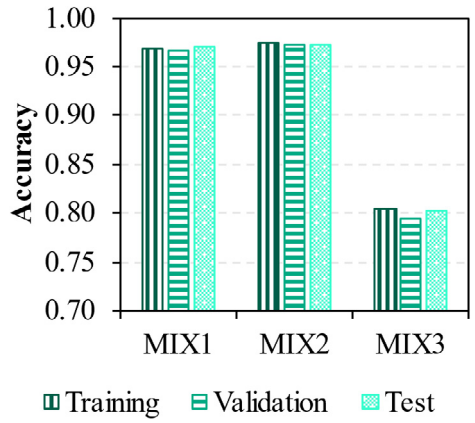


Fig. 12. Training, validation, and test accuracies of the SVM model for MIX1, MIX2, and MIX3.

models using nested CV. This methodology aimed to reduce overfitting, which was critical for the current application. Predictive models were developed for the ER classification of each fuel blend using three ML algorithms (LR, SVM, and ANN [MLP]). The performance of the predictive models was evaluated using a training–test split and nested CV within the training set by computing their accuracy, confusion matrices, and learning curves.

The model developed in this work allowed the prediction of the ER using a mean step size of 0.10, which in the case of the selected fuel mixtures, implied differences that would be difficult to be perceived by the human eye. The detection of slight changes in the combustion conditions allowed for the correction of deviated parameters, helping to optimize the processes and avoid the appearance of critical instabilities.

It was confirmed that the computed IFs used were affected by the ER class regardless of the fuel heating value. However, the subset of IFs with the highest variances depended on the fuel heating value. The subsets for the first and second mixtures shared three IFs: the standard deviation of the green and blue channels and the difference entropy of the green channel. The subsets of the three fuel blends shared only one IF, the centroid vertical coordinate of the blue channel. Moreover, the subsets of IF for all heating values included textural and geometrical features from the green and blue channels.

For the different fuel heating values, using different ML algorithms (LR, SVM, and ANN [MLP]) had no significant effect on accuracy. Nevertheless, the SVM models always provided the best results. Predictive models showed a relevant decrease in accuracy for the conditions with the lower fuel heating value (MIX3), reducing its value from 0.95–0.97 to 0.78. Considering SVM models, significant misclassifications for MIX3 occurred for consecutive ER classes with the smallest step sizes (0.02–0.03). However, for the conditions with a higher fuel heating value (MIX1), the predictive models obtained a higher accuracy when predicting consecutive ER classes with a step size of 0.03.

The SVM models showed a satisfactory fit to the data without underfitting or overfitting. The models achieved stability with only 25% of the flame images. Therefore, the test duration or frame rate could be reduced. Finally, the accuracy of the SVM models was measured again using a test set with previously unused flame images. The SVM models showed similar values to previous accuracies; thus, they performed well with unseen flame images.

The decreased accuracy reported for the fuel blend with the lower heating value could be used to propose future studies focused on monitoring fuel blends with low heating values to increase prediction accuracy. The current research sets the stage for automated monitoring of minor variations in the combustion of gaseous fuels.

Finally, the current research results, carried out at a laboratory level, would enable the development of these systems for their final implementation in industrial furnaces of the steel sector, where the mixtures used in this work present an alternative for fossil fuel substitution and, thus, for their decarbonization.

Funding

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 820771 (BAMBOO project) and 869939 (RETROFEED project).

CRediT authorship contribution statement

Pedro Compais: Conceptualisation, Methodology, Software, Formal Analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualisation. **Jorge Arroyo:** Conceptualisation, Methodology, Validation, Investigation, Resources, Writing – review & editing, Supervision, Project Administration. **Miguel Ángel Castán-Lascorz:** Software, Data curation, Writing – review & editing. **Jorge Barrio:** Software, Data curation, Writing – review & editing. **Antonia Gil:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jorge Arroyo reports financial support was provided by European Commission.

Data availability

The data that has been used is confidential.

References

Abdulaal, M.J., Casson, A.J., Gaydecki, P., 2018. Performance of nested vs. non-nested SVM cross-validation methods in visual BCI: Validation study. In: European Signal Processing Conference. Vol. 2018, pp. 1680–1684. http://dx.doi.org/10.23919/EUSIPCO.2018.8553102.

Abdurakipov, S.S., Gobyzov, O.A., Tokarev, M.P., Dulin, V.M., 2018. Combustion regime monitoring by flame imaging and machine learning. Optoelectron. Instrument. Data Process 54, 513–519. http://dx.doi.org/10.3103/S875669901805014X.

Aliramezani, M., Koch, C.R., Shahbakhti, M., 2022. Modeling, diagnostics, optimization, and control of internal combustion engines via modern machine learning techniques: A review and future directions. Prog. Energy Combust. Sci. 88, 100967. http://dx.doi.org/10.1016/j.pecs.2021.100967.

Bai, X., Lu, G., Hossain, M.M., Szuhánski, J., Daood, S.S., Nimmo, W., Yan, Y., Pourkashanian, M., 2017. Multi-mode combustion process monitoring on a pulverised fuel combustion test facility based on flame imaging and random weight network techniques. Fuel 202, 656–664. http://dx.doi.org/10.1016/j.fuel.2017.03.091.

Caillat, S., 2017. Burners in the steel industry: Utilization of by-product combustion gases in reheating furnaces and annealing lines. Energy Procedia 120, 20–27. http://dx.doi.org/10.1016/j.egypro.2017.07.152.

Cawley, G.C., Talbot, N.L.C., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. J. Mach. Learn. Res. 11, 2079–2107.

Cheng, Y., Huang, Y., Pang, B., Zhang, W., 2018. ThermalNet: A deep reinforcement learning-based combustion optimization system for coal-fired boiler. Eng. Appl. Artif. Intell. 74, 303–311. http://dx.doi.org/10.1016/j.engappai.2018.07.003.

Compais, P., Arroyo, J., González-Espinosa, A., Castán-Lascorz, M.Á., Gil, A., 2022a. Optical analysis of blast furnace gas combustion in a laboratory premixed burner. ACS Omega 7, 24498–24510. http://dx.doi.org/10.1021/acsomega.2c02103.

Compais, P., Arroyo, J., González-Espinosa, A., Gonzalo-Tirado, C., Castán-Lascorz, M.A., Barrio, J., Cuervo-Piñera, V., 2022b. Experimental analysis of blast furnace gas co-firing in a semi-industrial furnace using colour images. In: Proceedings of the 7th World Congress on Momentum, Heat and Mass Transfer, Apr. 2022, Paper No. CSP 117. http://dx.doi.org/10.11159/csp22.117.

Cuervo-Piñera, V., Cifrián-Riesgo, D., Battaglia, V., Fantuzzi, M., Della-Rocca, A., Ageno, M., Rensgard, A., Wang, C., Niska, J., Ekman, T., Rein, C., Nguyen, P.D., 2018. High Efficiency Low NOX BFG Based Combustion Systems in Steel Reheating Furnaces (HELNOx-BFG). final report, Publications Office, European Commission, Directorate-General for Research and Innovation.

Cuervo-Piñera, V., Cifrián-Riesgo, D., Nguyen, P.D., Battaglia, V., Fantuzzi, M., Della Rocca, A., Ageno, M., Rensgard, A., Wang, C., Niska, J., Ekman, T., Rein, C., Adler, W., 2017. Blast furnace gas based combustion systems in steel reheating furnaces. Energy Procedia 120, 357–364. http://dx.doi.org/10.1016/j.egypro.2017.07.215.

González-Cencerrado, A., Gil, A., Peña, B., 2013. Characterization of PF flames under different swirl conditions based on visualization systems. Fuel 113, 798–809. http://dx.doi.org/10.1016/j.fuel.2013.05.077.

González-Cencerrado, A., Peña, B., Gil, A., 2012. Coal flame characterization by means of digital image processing in a semi-industrial scale PF swirl burner. Appl. Energy 94, 375–384. http://dx.doi.org/10.1016/j.apenergy.2012.01.059.

González-Cencerrado, A., Peña, B., Gil, A., 2015. Experimental analysis of biomass co-firing flames in a pulverized fuel swirl burner using a CCD based visualization system. Fuel Process. Technol. 130, 299–310. http://dx.doi.org/10.1016/j.fuproc.2014.10.041.

González-Espinosa, A., Gil, A., Royo-Pascual, L., Nueno, A., Herce, C., 2020. Effects of hydrogen and primary air in a commercial partially-premixed atmospheric gas burner by means of optical and supervised machine learning techniques. Int. J. Hydrogen Energy 45, 31130–31150. http://dx.doi.org/10.1016/j.ijhydene.2020.08.045.

Han, Z., Hossain, M.M., Wang, Y., Li, J., Xu, C., 2020. Combustion stability monitoring through flame imaging and stacked sparse autoencoder based deep neural network. Appl. Energy 259, 114159. http://dx.doi.org/10.1016/j.apenergy.2019.114159.

Han, Z., Li, J., Zhang, B., Hossain, M.M., Xu, C., 2021. Prediction of combustion state through a semi-supervised learning model and flame imaging. Fuel 289, 119745. http://dx.doi.org/10.1016/j.fuel.2020.119745.

Hanuschkin, A., Zündorf, S., Schmidt, M., Welch, C., Schorr, J., Peters, S., Dreizler, A., Böhm, B., 2021. Investigation of cycle-to-cycle variations in a spark-ignition engine based on a machine learning analysis of the early flame kernel. Proc. Combust. Inst. 38, 5751–5759. http://dx.doi.org/10.1016/j.proci.2020.05.030.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. http://dx.doi.org/10.1109/TSMC.1973.4309314.

Huang, M., Xiao, Y., Zhang, Z., Shao, W., Xiong, Y., Liu, Y., Liu, Z., Lei, F., 2015. Effect of air/fuel nozzle arrangement on the MILD combustion of syngas. Appl. Therm. Eng. 87, 200–208. http://dx.doi.org/10.1016/j.applthermaleng.2015.04.076.

Ibargüengoytia, P.H., Delgadillo, M.A., García, U.A., Reyes, A., 2013. Viscosity virtual sensor to control combustion in fossil fuel power plants. Eng. Appl. Artif. Intell. 26, 2153–2163. http://dx.doi.org/10.1016/j.engappai.2013.05.004.

Jung, M.Y., Chang, J.H., Oh, M., Lee, C., 2023. Dynamic model and deep neural network-based surrogate model to predict dynamic behaviors and steady-state performance of solid propellant combustion. Combust. Flame 250, 112649. http://dx.doi.org/10.1016/j.combustflame.2023.112649.

Katzer, C., Babul, K., Klatt, M., Krautz, H.J., 2017. Quantitative and qualitative relationship between swirl burner operating conditions and pulverized coal flame length. Fuel Process. Technol. 156, 138–155. http://dx.doi.org/10.1016/j.fuproc.2016.10.013.

Lawal, M.O., 2021. Tomato detection based on modified YOLOv3 framework. Sci. Rep. 11, 1447. http://dx.doi.org/10.1038/s41598-021-81216-5.

Liu, Y., Xue, Q., Zuo, H., Yang, F., Peng, X., Wang, J., 2021. Effects of CO_2 and N_2 dilution on the combustion characteristics of H_2/CO mixture in a turbulent, partially premixed burner. ACS Omega 6, 15651–15662. http://dx.doi.org/10.1021/acsomega.1c00534.

Mathew, A.P., Asokan, A., Batri, K., Sivakumar, D., 2016. Comparative analysis of flame image features for combustion analysis. Indian J. Sci. Technol. 9, 1–11. http://dx.doi.org/10.17485/ijst/2016/v9i6/79904.

Matkvoic, F., Ivasic-Kos, M., Ribaric, S., 2022. A new approach to dominant motion pattern recognition at the macroscopic crowd level. Eng. Appl. Artif. Intell. 116, 105387. http://dx.doi.org/10.1016/j.engappai.2022.105387.

Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B.A., Del Rio-Chanona, E.A., Zhang, D., 2021. Machine learning for biochemical engineering: A review. Biochem. Eng. J. 172, 108054. http://dx.doi.org/10.1016/j.bej.2021.108054.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. SMC-9, 62–66. http://dx.doi.org/10.1109/TSMC.1979.4310076.

Park, M., Hur, J., Lee, W., 2022. Prediction of oil-fired boiler emissions with ensemble methods considering variable combustion air conditions. J. Clean. Prod. 375, 134094. http://dx.doi.org/10.1016/j.jclepro.2022.134094.

Quesada, D., Valverde, G., Larrañaga, P., Bielza, C., 2021. Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks. Eng. Appl. Artif. Intell. 103, 104301. http://dx.doi.org/10.1016/j.engappai.2021.104301.

Roy, A.M., 2022. Adaptive transfer learning-based multiscale feature fused deep convolutional neural network for EEG MI multiclassification in brain–computer interface. Eng. Appl. Artif. Intell. 116, 105347. http://dx.doi.org/10.1016/j.engappai.2022.105347.

Singer, G., Cohen, Y., 2021. A framework for smart control using machine-learning modeling for processes with closed-loop control in Industry 4.0. Eng. Appl. Artif. Intell. 102, 104236. http://dx.doi.org/10.1016/j.engappai.2021.104236.

Sreedhar, K., Panlal, B., 2012. Enhancement of images using morphological transformations. Int. J. Comp. Sci. Inf. Technol. 4, 33–50. http://dx.doi.org/10.5121/ijcsit.2012.4103.

Sun, D., Lu, G., Zhou, H., Li, X., Yan, Y., 2013. A simple index based quantitative assessment of flame stability. In: 2013 IEEE International Conference on Imaging Systems and Techniques. IST, Vol. 2013, pp. 190–193. http://dx.doi.org/10.1109/IST.2013.6729689.

Sun, D., Lu, G., Zhou, H., Yan, Y., Liu, S., 2015. Quantitative assessment of flame stability through image processing and spectral analysis. IEEE Trans. Instrum. Meas. 64, 3323–3333. http://dx.doi.org/10.1109/TIM.2015.2444262.

Vaish, R., Dwivedi, U.D., Tewari, S., Tripathi, S.M., 2021. Machine learning applications in power system fault diagnosis: Research advancements and perspectives. Eng. Appl. Artif. Intell. 106, 104504. http://dx.doi.org/10.1016/j.engappai.2021.104504.

Wainer, J., Cawley, G., 2018. Nested cross-validation when selecting classifiers is overzealous for most practical applications. http://dx.doi.org/10.48550/arXiv.1809.09446, arXiv:1809.09446.

Wang, Z., Zhao, X., Han, Z., Luo, L., Xiang, J., Zheng, S., Liu, G., Yu, M., Cui, Y., Shittu, S., Hu, M., 2021. Advanced big-data/machine-learning techniques for optimization and performance enhancement of the heat pipe technology – a review and prospective study. Appl. Energy 294, 116969. http://dx.doi.org/10.1016/j.apenergy.2021.116969.

Yang, G., He, Y., Li, X., Liu, H., Lan, T., 2022. Gabor-GLCM-based texture feature extraction using flame image to predict the O_2 content and NO_x. ACS Omega 7, 3889–3899. http://dx.doi.org/10.1021/acsomega.1c03397.

Zheng, W., Pang, L., Liu, Y., Xie, F., Zeng, W., 2021. Effects of methane addition on laminar flame characteristics of premixed blast furnace gas/air mixtures. Fuel 302, 121100. http://dx.doi.org/10.1016/j.fuel.2021.121100.

Zhu, J., Wang, Z., Li, R., Liu, S., Hua, Y., 2023. Experimental study and prediction model of combustion stability and combustion mode variation of burning methanol/biodiesel blends for diesel engines. Fuel 335, 127038. http://dx.doi.org/10.1016/j.fuel.2022.127038.