



Early fire detection using wavelet based features

Houda Harkat ^{a,b,*}, Hasmath Farhana Thariq Ahmed ^c, José M.P. Nascimento ^{a,d}, Alexandre Bernardino ^e

^a COPELABS, Universidade Lusófona de Humanidades e Tecnologias, 1749-024 Lisboa, Portugal

^b Instituto de Telecomunicações, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001, Lisboa, Portugal

^c School of Computer Science, University of Southampton Malaysia, 79100 Iskandar Puteri, Johor, Malaysia

^d Instituto Superior de Engenharia de Lisboa, IPL, Lisbon, Portugal

^e ISR - Instituto de Sistemas e Robotica, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

ARTICLE INFO

Keywords:

Wildfire

Wavelet-based features

Feature selection

Support Vector Machine (SVM)

Radial Basis Function (RBF)

ABSTRACT

In recent years, millions of hectares of vegetation worldwide have been destroyed by wildfires and forest fires. Computer vision-based fire classification, which separates pixels from image or video datasets into fire and non-fire categories, has gained more attention recently due to technological innovations. Fire pixels in an image or video can be classified using either a deep learning strategy or a traditional machine learning approach. Deep learning algorithms could process enormous volumes of data, but their training model performance is constrained because they fail to consider the differences in complexity between training samples. Moreover, it is not obvious to train a deep learning model without a dedicated GPU unit.

Similarly, deep learning techniques that have a scarcity of training data and insufficient features exhibit poor performance in intricate real-world fire situations. Consequently, to categorize fire and non-fire pixels from the processed photos from the publicly accessible datasets, Corsican and FLAME, as well as the aerial private dataset Firefront_Gestosa, the current study uses a lightweight technique based on SVM and a refined set of features.

The present research implemented a novel framework for fire detection and classification from a variety of RGB and Infra-red images acquired during real missions, addressing the significant requirement for swift and accurate recognition of diverse types of flames, ranging from wildfires to industrial and domestic fires. The framework employs wavelet decomposition-based features, including wavelet length, standard deviation, variance, energy, and Shannon's entropy, extracted through a sliding window sampling method within a machine learning approach. It should be noted that managing multidimensional data to train a model is difficult in machine learning applications. This issue is solved adopting a feature selection approach, which eliminates redundant or unnecessary data that affects the functionality of the model. Thus, to enhance model performance, feature selection using ranking algorithms based on theoretical mutual information is applied in combination to the Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel.

Extensive experiments demonstrate exceptional results, notably with Haar wavelets, achieving an impressive overall accuracy of 99.43% and remarkable performance across specificity, precision, recall, F-measure, and G-mean metrics. Thus, the present study showcases the potential of advanced image processing techniques to significantly advance fire detection and classification, thereby contributing to fire prevention, management, and research in various contexts.

1. Introduction

Establishing precise and reliable forest fire detection and surveillance systems is essential to mitigating the adverse consequences of forest fires. Consequently, numerous forest fire surveillance systems

implement diverse technologies, including drones, satellite images, and ground sensors, to promptly identify, analyze, and respond to forest fire incidents. The implementation of these sensors has resulted in substantial improvements in the detection of forest fires. Fire image detection and classification is a subfield of computer vision that aims to

* Corresponding author.

E-mail addresses: houda.harkat@usmba.ac.ma (H. Harkat), H.F.Thariq-Ahmed@soton.ac.uk (H.F.T. Ahmed), jose.nascimento@isel.pt (J.M.P. Nascimento), alex@isr.tecnico.ulisboa.pt (A. Bernardino).

automatically identify and classify images of fires into different classes, such as wildfire [22], industrial fire [62], and house fire [59].

Fire detection is critical as it can help with fire prevention and management by providing timely and accurate information about the location and severity of fires. In the case of wildfires, for example, fire image classification can be used to monitor the spread of the fire, predict its trajectory, and informed decision-making for resource allocation and evacuation planning [8]. In industrial settings, fire image classification can aid in the identification of potential fire hazards and the activation of early warning systems, thereby preventing significant damage and loss of life [1].

Traditional machine learning and deep learning approaches have been widely adopted for fire detection and classification (Cao, Tang, Wu, & Lu, 2021; C. Yang, Pan, Cao, & Lu, 2022). Traditional machine learning-based approaches rely solely on classical signal processing techniques for feature extraction combined with computationally efficient learning-based classification models [12,18,22]. Deep learning-based approaches rely on signal processing algorithms in combination with deep learning detection models [23,24,31], namely, You Only Look Once (YOLO) [5,33], R-CNN [60] and its variants. It is worth noting that the state-of-the-art focuses widely on deep learning-based approaches rather than traditional approaches.

Machine learning algorithms can be trained on a dataset of images with fire and non-fire pixels to identify the unique patterns and features that distinguish the fire and non-fire pixels from the image. State-of-the-art machine learning-based approaches widely adopted logistics regression, Support Vector Machine (SVM), and fuzzy logic for accurate classification of fire images. The Sequential Minimal Optimization for Regression (SMOreg) Choudhary et al. [13] model is designed to analyze images captured by drones to detect forest fires. With an accuracy rate of 92.72 %, this model can accurately identify the presence of forest fires by being trained on a large dataset of images that includes both positive and negative examples. Furthermore, logistic regression, in combination with vanilla CNN, is adopted for analyzing and detecting the fire occurrences in the multispectral images [3].

In a recent study, a Preferred Vector Machine (PreVM) [65], based on SVM, is implemented to enhance the precision of pixel-level fire detection and decrease error rates. The kernel-based variant improves the real-time detection performance and model training efficiency, –PreVM. Moreover, in another study, image data was initially used to classify different forest categories. However, a more detailed analysis was conducted using the Artificial Bee Colony-Adaptive Neuro-Fuzzy Inference System (ABC-ANFIS) model, which examined a variety of environmental and potentially image-derived factors to assess the risk of forest fires [15]. In fire detection and classification, machine learning algorithms can be trained on a fire and non-fire images dataset to identify the distinguishing patterns and characteristics of fire images [7]. In another recent work, the DeepFire dataset was created with real-world forest fire and non-fire images and evaluated the fire detection performance comparatively with multiple machine learning algorithms, such as –nearest neighbors, random forest, naive Bayes, SVM, and logistic regression [30]. The simulation results indicate classification accuracy, with a mean accuracy of 95 % and precision and recall of 95.7 % and 94.2 %, respectively.

Deep learning algorithms have proven better results in the study of fire detection. State-of-the-art deep learning approaches implement various architectures, namely Inception [29], Resnet [14], Inception-Resnet [43], RNN [17], MobileNetV2 [43], and DenseNet [39], applied to a wide range of image datasets. Deep learning algorithms have proven better results in the study of fire detection as they can analyze enormous amounts of data and autonomously learn image features to classify forest fire images from non-fire images [8,21,35]. However, one of the most significant challenges associated with state-of-the-art deep learning-based approaches is the requirement for an extensive variety of high-quality training data that accurately replicates various forest fire incident scenarios. To capture this collection of

complex feature representations, a deep learning network must have a substantial quantity of diverse training data [45].

Irrespective of the choice of learning approaches, the performance of the learning algorithms relies on features extracted from the image dataset, as the quality of the extracted features impacts overall model accuracy. In a recent study, a Wavelet-CNN method [24] is implemented with 2D Haar transform, which extracts spectral features as input to the CNN, and the performance is evaluated with MobileNet v2. MobileNet v2 reported improved fire detection performance with fewer false alarms; however, not all false positives were eliminated. A study based on elastic net [67], for forest fire smoke detection extracts quasi-dynamic features, adopting a dual-tree complex wavelet transform for decomposing images in sub-bands [10]. Thus, capturing the smoke in different orientations [63] The proposed approach is less time-consuming than conventional deep approaches. However, deep learning approaches over a different dataset achieve similar or better results. However, no direct performance comparison was made with other techniques for the same set of data.

The underlying idea is that in some real-time scenes, it is essential to have a quick preliminary detection tool to detect fire for immediate intervention. Rather than adopting deep learning-based approaches that require complex hardware, it is preferable to have a quick decision model that could be run on a simple mobile unit and easily converted to an FPGA unit to allow real-time processing. Nonetheless, with the vast amount of data, the training process becomes very time-consuming and requires adequate high-performance Graphic Processor Unit (GPU) based stations. It is worth noting that machine learning algorithms do not necessitate computationally expensive hardware for training.

Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbour (k-NN) are the widely used machine learning algorithms for this task that particularly do not require any high-capacity treatment unit to perform the training task. The present research opted to extract the wavelet coefficient from signals (the Radon transform of the pictures over thirty-seven projection angles) to obtain a deeper insight into texture details.. Furthermore, a diversified feature vector is extracted to reduce the dimensionality. The dataset adopted in the present study is diversified as well, as it contains pictures from aerial views and pictures with small-scale flames. Despite the complexity of detection over aerial images, the framework adopted in the present study draws an overall accuracy between 98 % and 99 %.

The Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel receives the chosen subset of feature vector as input and clusters the images into fire and non-fire classes. The model's performance is evaluated using its overall accuracy, specificity, precision, recall, F-measure, and G-mean. The tests were performed over two publicly available datasets, the Corsican [56] and FLAME [49] datasets, in addition to a private dataset Firefront_Gestosa collected by the Firefront project team during the Gestosa mission.

The present research is motivated by the necessity for a rapid, efficient fire detection tool capable of functioning on lightweight, real-time systems, such as mobile or FPGA units, without relying on specialized hardware. In contrast to deep learning models that require large datasets, GPUs, and increased processing times, machine learning algorithms utilizing manual feature extraction may offer more rapid, less expensive computational strategies, making them more appropriate for prompt intervention in real-time fire detection situations. The main contributions of this study are as follows:

- (1) Once the dynamic properties of the fire picture data have been correctly annotated and labeled (a set of private data is manually labeled), these labeled data serve as the basis for flame recognition. The dynamic features derived from wavelet analysis, data fitting, and frequency analysis can precisely describe the instate variations in various fire conditions. The purpose of the feature extraction schema is to gather the most significant associated

- texture and color data into a single, massive feature vector, which is then filtered.
- (2) A lightweight framework based on a set of limited features extracted utilizing wavelet decomposition, namely: wavelet length, standard deviation, variance, energy, and Shannon's entropy. The features are extracted using a sliding window sampling approach and then concatenated to form a relatively huge feature vector. To eliminate potential redundancy of information, ranking algorithms based on theoretical mutual information are implemented to choose the best set of features that allows optimal training of the model.
- (3) In-depth assessments are designed and carried out to demonstrate the efficacy of the suggested algorithm in comparison to other algorithms currently in use for identifying fires/flames. Several experiments are performed to optimize the model performance to choose preliminary parameters like mother wavelet, decomposition level, and adequate feature selection technique.

The remainder of the paper is structured as follows: Section 2 gives an insight into early fire detection approaches proposed through the literature, Section 3 describes the dataset used and presents the framework methods, detailing the wavelet-based feature extraction technique, the theoretical information-based feature ranking approaches deployed, and the classification strategy adopted. Moreover, the process of dataset construction, preparation, and processing is explained in detail. Section 4 covers the prototype model and the automatic fire detection results evaluated in terms of variation of different related parameters. Section 5 summarizes and concludes the present work.

2. Methods and dataset

The proposed framework deployed in this study is illustrated in Fig. 1. Firstly, the dataset is fused and preprocessed appropriately. The dataset in this work is a fusion of 2 public datasets and one private in addition to some sun pictures. Experts in the field labeled the private dataset, and the two public sets are obtained with their corresponding ground truths. The pictures were patched using a simple algorithm to localize the flames' position precisely. More detailed information about the procedure and processing are presented in subsequent paragraphs.

Table 1. Represents the pseudo-code of the model training algorithm. Once the data has been appropriately partitioned, a Radon transform is

applied to every picture to transform the data from 2D to 1D signals. Relying on a sliding window sampling approach, the signals are sampled with a predefined sampling frequency, and a set of wavelet coefficients is extracted from every portion. Five features are calculated based on those coefficients for each channel. For a five-level decomposition, thirty features are obtained per window from each channel. The features are concatenated to a unique feature vector reduced afterward using ranking approaches. The objective is to reduce information redundancy to expedite the training phase. The parameters of the adopted SVM classifier are adequately optimized to enhance the results. It is imperative to note that the appropriate selection of the mother wavelet and decomposition level is crucial.

The subsequent section presents further details and theoretical aspects adopted in the present study. Post-training, the SVM model is utilized to detect fire regions in higher-dimensional images using a sliding window approach. The results shown in Fig. 1 illustrate the object detection outcomes, including the corresponding bounding boxes, in the final phase.

2.1. The dataset

2.1.1. Original dataset

The datasets Corsican [56] and FLAME [49] are both openly accessible and were used in the present work to separate fire from non-fire regions in corresponding pictures. Furthermore, the analysis in the present study is conducted using the images that the FireFront project team had acquired during the Gestosa mission.

The Corsican dataset is a collection of labeled images used to detect fires. The dataset was created using images captured by different camera types from around the world. The dataset contains more than 2000 labeled images. The dataset contains both flame and non-flame (background) shots in the visible and near-infrared (NIR) bands, and professionals labeled the pictures. Various multimodal images captured with a JAI AD-080GE camera compensate for the dataset. This kind of camera can capture images in the visible and NIR bands using the same synchronized optics. Only 1775 images out of the whole image dataset with mixed flame color and profound quality factors (textures and illumination) were chosen to train the fire detection model. Each picture is paired with a corresponding segmentation mask generated via a homography matrix transform-based approach for the Corsican dataset.

The FLAME dataset is another publicly available dataset acquired by

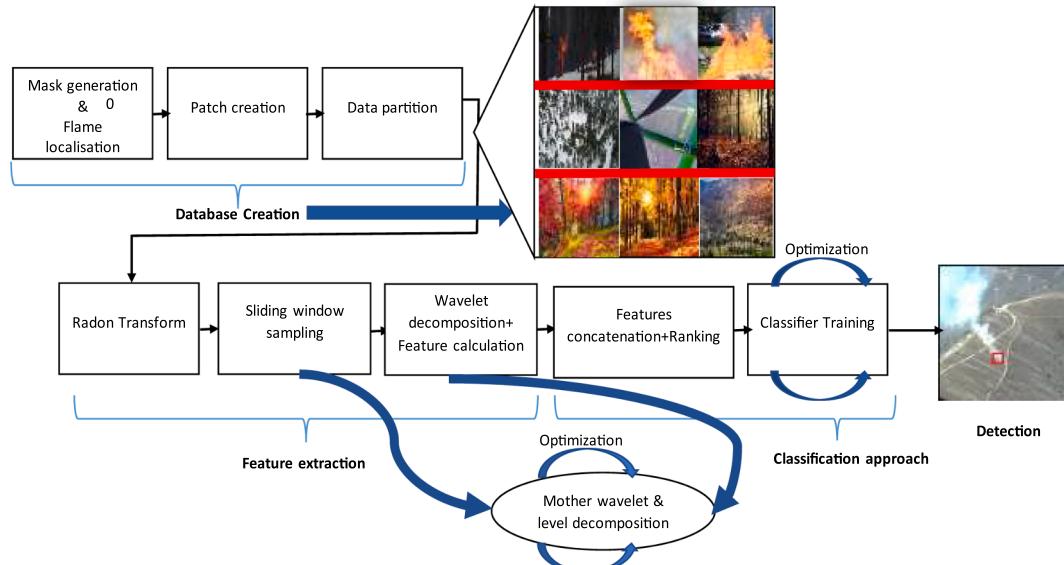


Fig. 1. The current introduced framework for fire detection based on wavelet decomposition.

Table 1

Model training algorithm.

Input	Imdb.images.data > 8036 samples (every sample is a matrix 300x300)	
Features extraction	<pre> For i = 1: 8036 0=0; R ← radon(imdb.images.data(:,i, 0); cmat ← getmswfeat(R,128,512,512); ho1 = [cmat']; ho1 = 0; For θ =10:10:360; R ← radon(imdb.images.data(:,i, θ); cmat ← getmswfeat(R,128,512,512); ho1 = [ho1;cmat']; End features(:,i) = [ho1]; End </pre>	> Compute Radon transform for the corresponding 0. > Compute the frequency features for the corresponding radon projection.
SVM training	<pre> cv = cvpartition(8036,'kfold',5); feat = normalize(feat); selectedIndices = feat('ICAP',80,feat,label); feattt = feat(:,selectedIndices); cg = [0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300]; For c = 1:1:11 For g = 11:-1:1 Log2c = cg(c); Log2g = cg(g); MM = zeros(1,7); For k = 1:5 testIdx = find(test(cv,k)); trainIdx = find(training(cv,k)); heart_scale_inst = feat(trainIdx,:); heart_scale_inst = normalize(heart_scale_inst); heart_scale_label = label(trainIdx); For i = 1:5 cmd = ['-s 0 -t 2 -c ',num2str(log2c),' -g ',num2str(log2g),' -h 0']; model = svmtrain(heart_scale_label, heart_scale_inst,'cmd'); heart_scale_inst = feat(:,i); heart_scale_label = label; [predict_label, accuracyov, prob_estimates] = svmpredict(label, feat, model); End End End End </pre>	> Data partition. > Normalize feature vector. > Select features with ICAP. >Used values for training SVM for optimization purpose. >Data partition: training and test. >Applied 5-fold cross validation and took the average.
Output	Model	>Trained model

drones in the northern part of Arizona. The data compilation includes fire videos and photos taken during a controlled slash-pile burn. The chosen films were recorded with a Zenmuse X4S and a Phantom 3 camera with a 640×512 pixels resolution and a 30-fps frame rate.

In the current research, it was recommended to use 2003 photos that had already been tagged with each image's relevant ground truth, with some converted as frames to train the model. Section 3.1.3 will present more detailed information about the transformations and preprocessing done for every set of data.

Another unlabeled, private aerial dataset used in the current study is Firefront Gestosa. The FireFront project team (<https://firefront.pt/>) collected this dataset during the Gestosa mission. Researchers will presumably have access to it in the near future. This dataset consists of five recorded videos, each lasting between two and three minutes. Please refer to Fig. 2 for more information on each video.

Instead of being a videotape of the mission, the sixth and ultimate video is a mixture of various viewpoints from the first five. This dataset also poses a substantial problem because it includes photos that have been covered by smoke, making it challenging to isolate the flame from the surrounding area. The present study produced 238 fire frames from already-existing films for the experimental work. The fire images were manually labeled using MATLAB Image Labeler.

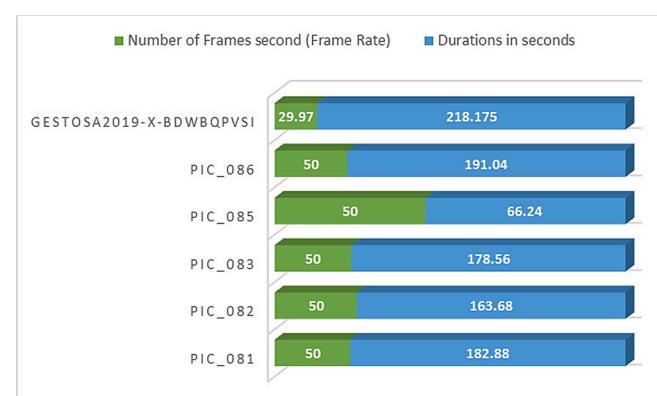


Fig. 2. Statistics about Firefront_Gestosa dataset: Frame rate and duration of each video.

2.1.2. Final dataset

The present study converts the original image data into a 300x300 grayscale patch. A basic patching technique was adopted for zooming the presence of fire and its position from Corsican and FLAME datasets.

Thus, the required data for experimental work in the present research is compressed (the algorithm is detailed in the next paragraph). The algorithm determines the positions of the fire pixels and flames based on the ground truths. Nonetheless, adopting such a technique seems impractical for Firefront_Gestosa due to the presence of a limited number of fire pixels in the images. Fig. 3. illustrates the dataset's patches.

The dataset used in this study's statistics are summarized in Fig. 4. The Corsican dataset and Flame dataset consist of 1775 fire pictures and 4845 (2003 fire and 2842 non-fire) shots, respectively. Whereas there are 1351 images in Firefront_Gestosa, with 238 fire and 1113 non-fire images. Furthermore, 65 non-fire images were included from "sun images." the present study performed experiments with the fused dataset of 8036 images, of which there are 4016 fire and 4020 nonfire shots. To minimize the number of false positives (FP) in the classification step, "sun images" are utilized to mimic high-intensity images (natural sunrise and set phenomena, and some high-brightness wild pictures).

Aerial datasets also contain images of fire recorded from very high altitudes, giving the viewer a detached view of the flame. As a result, clustering appears more complicated than it actually is. In addition, the pictures were patched with a resolution of 300x300 to cover all of the flame pixels, eliminating intermittent data processing while retaining crucial information.

2.2. Preprocessing of fire image dataset

The three data sets had gone through a preprocessing phase to improve the image quality and make them easier to store. The radon transform was then applied to reduce their dimension.

2.2.1. Preprocessing algorithm

The newly collected images from the Firefront Gestosa data were manually tagged with pixel markers according to human eyesight using "MATLAB ImageLabeler." The labels were then converted to binary pictures after being normalized. The preprocessing procedures used in the current study are outlined in the chart of Fig. 5. The mask matrix refers to ground truth images for every instance. Prior to converting them to grayscale images, the indexed ground truth images underwent normalization.

In addition, the images were once more converted to index values using an inverted colormap technique [55]. The algorithm quantizes the color map [55] into 25 unique nuance degrees for each color value. Subsequently, every pixel in the grayscale picture has its nearest nuance localized from the quantized colormap.

The main objective of the aforementioned image conversion processes, grayscale to indexed and indexed to grayscale, is to create images that use a colormap to simplify storage/treatment later. The images are then appropriately stored after being binarized using a specified threshold. The same preprocessing techniques were used on the FLAME dataset, which had previously been labeled. Prior to training the models,

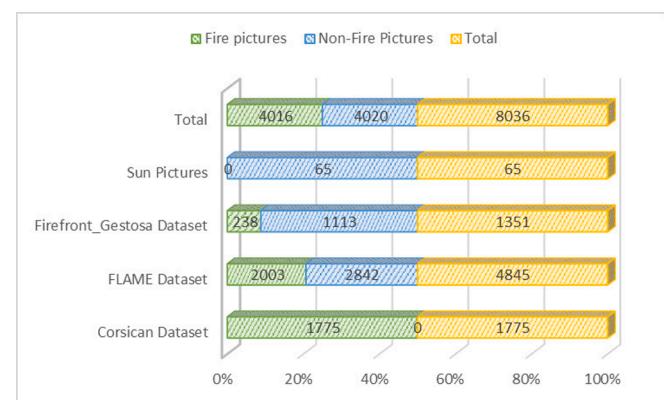


Fig. 4. The final dataset's distribution: the positive and negative contributions made by each type of data.

it was necessary to convert infrared images of Corsican data to RGB indexing format by merely replicating the red channel's content over that of the green and blue channels.

2.2.2. Radon transform

In the first stage, the Radon transform is applied to the pictures of the selected dataset. The total of each pixel's Radon transformations makes up an image's Radon transform [53]. The pixels in a picture are divided into four subpixels using this technique, and each subpixel is projected separately. Each subpixel's contribution is proportionally split between the two closest bins according to the gap between the projected position and the image's bin center. If the subpixel projection hits the bin center, the axis bin computes the total subpixel value or one-fourth of the pixel value. As input parameters for the Radon transform, a step size of 100 and a projection angle ranging from $\theta = 00$ to 3600 are used. As a result, 37 angles are formed for each image. The Radon transform, R, when applied to an image I with an angle Θ , is represented as follows:

$$R = \text{radon}(I, \Theta) \quad (1)$$

2.3. Wavelet-based feature extraction

Every radon projection is decomposed with discrete wavelet transform using a sliding window approach. Then, a set of five features is extracted per window from each channel. The procedure will be thoroughly explained in the following section.

2.3.1. Wavelet decomposition

The input or original signal exhibits highly correlated features compared to the sample size, resulting in model overfitting with the training data. Therefore, extracting features from the original input



Fig. 3. Sample of the final datasets of patches: (a) samples of positives, and (b) negatives.

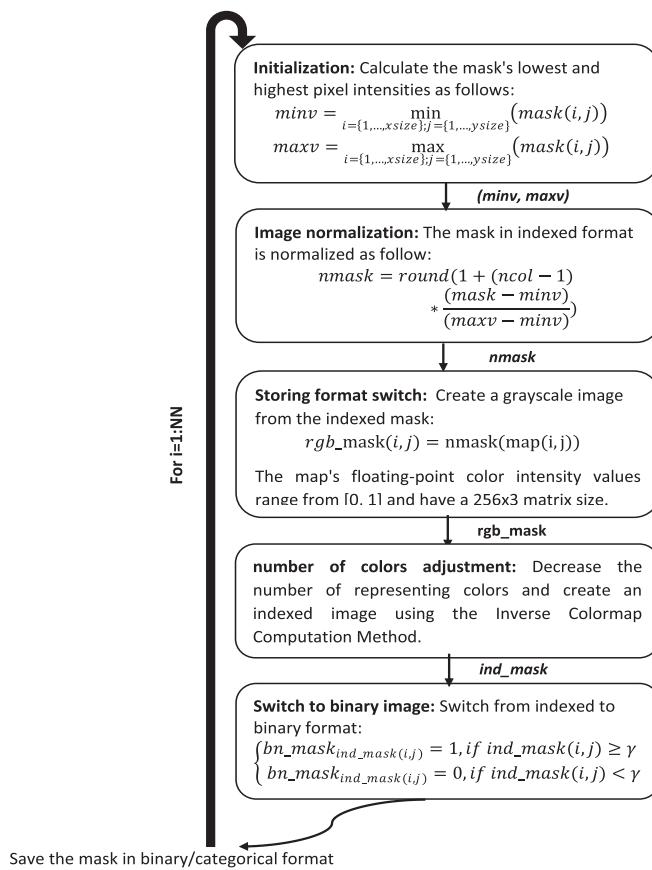


Fig. 5. Preprocessing algorithm: xsize and ysize represent the number of rows and columns of the ground truths, while ncol is the number of columns of the gray map. γ is the adopted threshold for binarization purposes, the standard value is 0.5. NN is the number of pictures to process.

signal is a fundamental step, as it greatly influences the model's performance.

Wavelet analysis involves converting a signal from a time to a time-frequency field. The analysis could be performed using continuous or discrete versions. At high frequencies, the Wavelet transform grants a decent time and alarming frequency resolution, whereas, at low frequencies, it has an acceptable frequency and good time resolution Peng et al. [41]. As a result of its ability to simultaneously capture time and frequency elements, it is employed to identify intermittent characteristics. For the signal type to be analyzed in this study, the Continuous Wavelet Transform (CWT) isn't an appropriate choice since it is characterized by the redundancy aspect [42]. The discrete wavelet transform, which was created to evaluate signals with a narrower range of scales and a precise number of transformations at each scale [52], can be equally efficient as the continuous version with the same accuracy. DWT is adopted in the present work, as it can efficiently distinctly signal details and remove noise with a low distortion rate compared to other transform techniques. Furthermore, DWT adaptively disassembles the time-frequency patterns of a signal, utilizing fewer wavelet coefficients for extracting fewer features.

Within discrete wavelet, the signals are divided into a hierarchy of detail and approximations at a well-predefined level, as characterized by the following equation:

$$R(t) = \sum_{k=1}^{k=J} D_k(t) + A_k(t) \quad (2)$$

where $R(t)$ is the original signal to be decomposed using wavelet at the time, $D_k(t)$ stands for the wavelet detail, and $A_k(t)$ for the corresponding wavelet approximation at the j^{th} level.

DWT serves as the foundation for wavelet packet analysis [4,36], a generalization of the wavelet decomposition concept with more analysis tools. Each transferred wavelet detail component is further divided into its own approximation and detail components in the hierarchy. A group of standard wavelet functions are linearly integrated to form a wavelet packet.

Wavelet packets inherit the orthonormality and time-frequency localization properties of their associated wavelet functions. In this study, the Radon signals have been sampled and then pre-processed using wavelet packets. The modulation, scale, and translation parameters are represented by the three integer indices α , β , and γ that make up a wavelet packet, as shown in equation (3). The recursive formula below is used to determine the wavelet functions $\psi^{\alpha}(t)$ [42]:

$$\left\{ \begin{array}{l} \psi_{\alpha,\beta}^{\gamma}(t) = 2^{j/2}\psi^{\gamma}(2^j t - \beta), \gamma = 1, 2, 3, \dots \\ \psi^{2\alpha}(t) = \sqrt{2} \sum_{-\infty}^{\infty} h(\beta)\psi^{\gamma}(2t - \beta) \\ \psi^{2\alpha+1}(t) = \sqrt{2} \sum_{-\infty}^{\infty} g(\beta)\psi^{\gamma}(2t - \beta) \end{array} \right. \quad (3)$$

$h(\beta)$ and $g(\beta)$ designate the corresponding low and high-pass filter bank structure [58].

Assuming that the wavelet packet functions meet the following orthogonality condition:

$$\left\{ \begin{array}{l} R(t) = \sum_{\gamma=1}^{2\alpha} R_{\alpha}^{\gamma}(t) \\ R_{\alpha}^{\gamma}(t) = \sum_{n=-\infty}^{\infty} c_{\alpha,\beta}^{\gamma}(t)\psi_{\alpha,\beta}^{\gamma}(t) \\ c_{\alpha,\beta}^{\gamma}(t) = \int_{-\infty}^{\infty} R(t)\psi_{\alpha,\beta}^{\gamma}(t)dt \end{array} \right. \quad (4)$$

$\psi_{\alpha,\beta}^{\gamma}(t)$ stands for the wavelet packet function for scale j , modulation α , and translation β .

The present study applied Discrete wavelet transforms (DWT) over the 1D radon-transformed signals, setting the sample window size to 512.

Five levels of wavelet decomposition are performed within the Symlet 8 mother wavelet [51] as initial experiments, different configurations were tested later. Per window, five features are extracted for each tree (the given mathematical formulation is resumed in Table 2): Waveform length, Standard deviation, Variance, Energy, and Shannon's entropy. Consequently, in the case of five-level decomposition, there are 30 features per projection angle. Thus, with 37 projection angles, a feature vector of 1100 is obtained.

In Table 2. N stands for the length of waveform, \bar{X} the mean value and X_i the i^{th} value.

2.3.2. Wavelet forms

The present study applied different mother wavelet types to assess the quality of extracted corresponding features.

The first and most basic orthonormal wavelet base is the Haar wavelet [54]. Conceptually straightforward, memory-efficient, precisely reversible, and computationally inexpensive, the Haar wavelet lacks the

Table 2
The set chosen features are calculated from the wavelet coefficients.

Feature	Mathematical formulation
Waveform length	$n(X) = N$
Standard deviation	$\rho(X) = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$
Variance	$\rho^2(X) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$
Energy	$\Psi(X^2) = - \sum_{i=1}^N X_i^2$
Shannon's entropy	$E(X^2) = - \sum_{i=1}^N X_i^2 \log_2 X_i^2$

edge effects present in other wavelets. The changes between neighboring pixel pairs are the only ones reflected by the Haar transform, which lacks overlapping windows. It computes pairwise averages and differences using two scaling and wavelet function coefficients.

Because of its capacity for orthogonal and compact support, the Daubechies wavelet [44] family is the most often used wavelet family for texture feature analysis. Because the Daubechies wavelet employs overlapping windows, all variations in pixel intensities are reflected in the results. Daubechies is smoother than the Haar wavelet because it averages over a more significant number of pixels. There are four wavelet and scaling coefficients in the Daubechies Db4 transform. Except for their higher symmetry, Symlets wavelets [51] resemble Daubechies wavelets in most respects. Symlets also exhibit the most vanishing moments and the most minor asymmetry among other wavelets. Since the total computation of the scaling function coefficients equals one, it averages over a set of four neighboring pixels.

The Daubechies wavelet served as the source for Coiflets [51,61]. It uses more overlapped windows and has an even higher computational overhead. Increases in pixel averaging and differencing result in a smoother wavelet and enhanced capabilities in various image-processing techniques (such as de-noising images) since it employs six scaling and wavelet function coefficients. The filter has the same structure as the Daubechies and Haar filters. With six neighboring pixels instead of two, it uses the same format to calculate differences and averages. The mirror filters are also used by the Coiflet wavelet [47].

The Daubechies wavelet families in the face-off Haar wavelet are more intricate and typically have a larger processing cost. The Haar wavelet outperforms Daubechies wavelets in some applications. It all depends on the processing goals and the analysis affinity to accomplish.

2.4. Feature ranking algorithm

The selection of a feature ranking algorithm and the quantity of features to be employed is critical. The theoretical elements of the various feature ranking approaches utilized in the present work are discussed in the subsequent paragraph.

The primary argument for adding a feature selection phase is that employing the entire set of derived features has a detrimental effect on computing capability. The inappropriate noise is filtered by choosing a portion of the most relevant feature inputs. Furthermore, removing information redundancies is important to allow fast and accurate training.

One of the measures used to assess the prominence of attributes is mutual information, which is founded on data theory. The most used algorithms are: Mutual Information Feature Selection (MIFS) [6], Min-Redundancy Max-Relevance (mRMR) [38], Peng et al. [40], Conditional Infomax Feature Extraction (CIFE) [32], Joint Mutual Information (JMI) [64], Conditional Mutual Information Maximization (CMIM) [16], Double Input Symmetrical Relevance (DISR) [37], Interaction Capping (ICAP) [26], and Conditional Redundancy (CONDRED) [9].

These information-theoretic methods rate each attribute according to how relevant it is to the goal variable but rather how redundant it is compared to other already chosen features. The algorithms choose the least redundant features with other features, and have the most significant relevance to the target. Until the requisite number of features is chosen, each feature is selected individually. The ideal selection process involves considering every conceivable combination of the n features, that is resource intensive. Thus, the process is sequential and demanding. The general conception of the algorithms is given in Table 3.

The mutual information-based evaluation formula differs from one algorithm to other. Table 4. shows the corresponding formula for every algorithm. P represents the probability function, MI is the mutual information, ν and a relative coefficient. D represents Kullback-Leibler divergence or relative entropy, which is a simple metric in information theory that quantifies how widely one distribution differs from another.

Table 3

Mutual information feature selection algorithm general concept.

Input	feature matrix, target data, required number of features (RN)
Processing	For n = 1: (RN) For each feature that has not been chosen yet: F1 ← Compute mutual information with each of the previously selected features (F1). Compute mutual information with the target variables. F2 ← Determine the mean values from the preceding steps (F2). Give the grade S = F1-F2. Add the attribute with the highest rating to the feature preference list. End End
Output	indices of picked features

Table 4

Mutual information feature selection algorithm.

Algorithm	Mutual information-based evaluation formula
MIFS	$MI(X_n, Y) - \nu \sum_{i=1}^{n-1} MI(X_n, Y_i)$
mRMR	$MI(X_n, Y) + \sum_{i=1}^{n-1} MI(X_n, Y_i)$
CIFE	$MI(X_n, Y) - \sum_{i=1}^{n-1} MI(X_n, Y_i) + \sum_{i=1}^{n-1} MI(X_n, Y_i Y)$
JMI	$\sum_{i=1}^{n-1} MI(X_n X_i, Y)$
CMIM	$MIN_i[MI(X_n, Y X_i)]$
DISR	$DISR(X_n, Y) = \left(\frac{1}{ X_n } \right) * \sum_{i=1}^n \frac{ X_n - D(Y X_n)}{D(Y)}$
ICAP	$ICAP(X_n, Y) = \frac{(P(X_n, Y) - P(X_n, Y'))}{\max(P(X_n, Y) - P(X_n, Y'))}$
CONDRED	$H(X_n, Y) = - \sum P(X_n Y) \log P(X_n Y)$

The computation of redundancy in relation to other attributes is what that make difference between each case. Nonetheless, each algorithm possesses varying degrees of advantages in comparison to the other, please refer to Fig. 6.

2.5. Classification Approach

The performance of the feature extraction technique was primarily evaluated using an SVM classifier and various other measures (apart from overall accuracy) that provide better vision.

The present research implemented a framework based on a limited number of radon transformed features derived from adopting wavelet decomposition, notably wavelet length, standard deviation, variance, energy, and Shannon's entropy. A sliding window sampling approach is used to extract the features, which are then concatenated to generate a relatively large feature vector. Ranking algorithms based on theoretical mutual information are used to select the best feature inputs for optimal model training by adopting SVM with the RBF kernel to reduce potential information redundancy.

The results reported based on extensive experimentation show significant performance with increasing feature numbers, specifically 80 MIFS and ICAP feature inputs. Thus, further experiments were performed with the MIFS and ICAP features inputs in combination with the Symlet series (3,5,8), Daubechies series (2,4,6), Haar and Coiflet series (2,4) considering overall accuracy, specificity, precision, recall, F-measure and G-mean as the evaluation metrics. However, Haar reported better detection performance with an overall accuracy of 99.43 %, with specificity, precision, recall, F-measure, and G-mean of 99.38 %, 99.38 %, 99.48 %, 99.43 %, and 99.43 %, respectively.

2.5.1. SVM classifier

SVM is a supervised machine learning algorithm adopted mainly for classification problems and performing regression analysis. Radial Basis

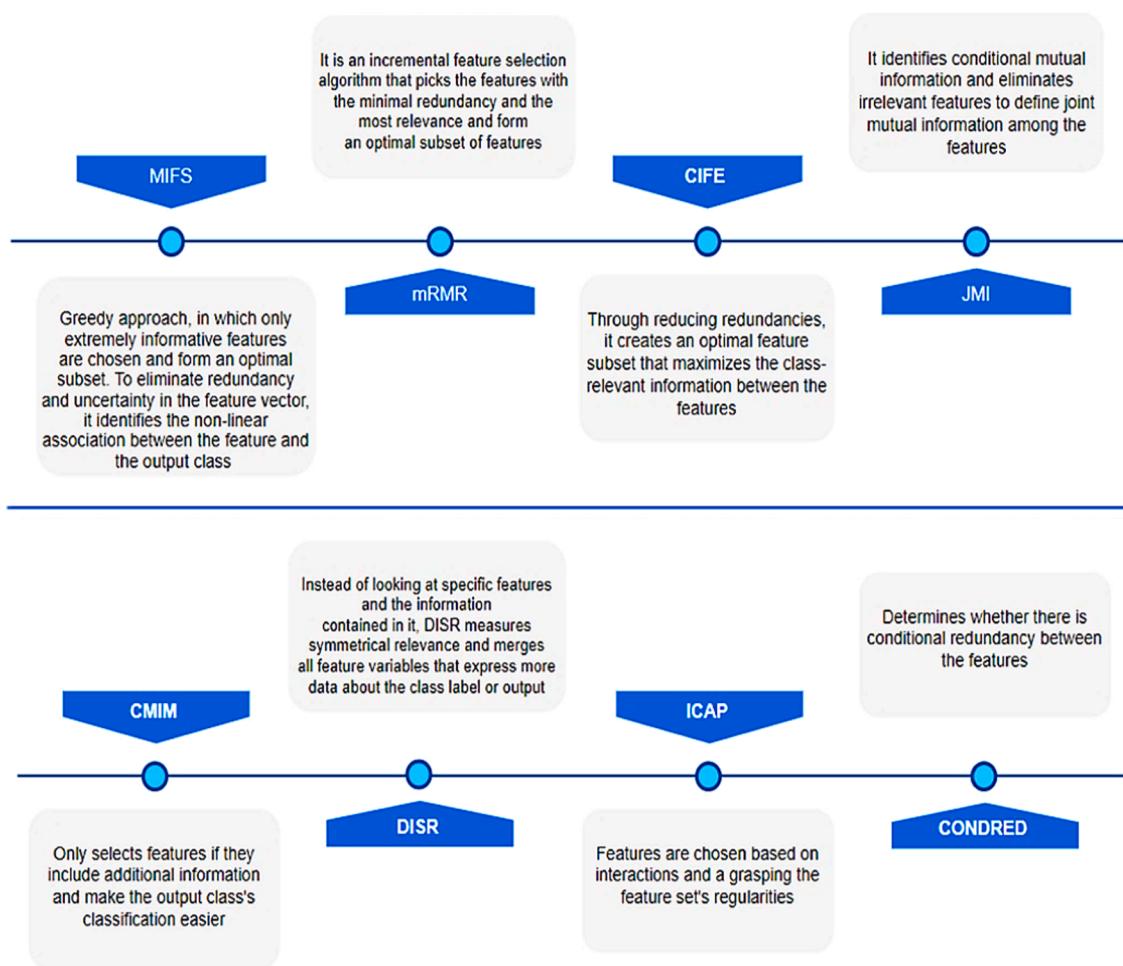


Fig. 6. Feature selection –MIFS, mRMR, CIFE, JMI, CMIM, DISR, ICAP, and CONDRED.

Function (RBF) is one of the most frequently used kernels for SVM. The RBF kernel is a nonlinear kernel that maps the input space to a higher-dimensional feature space that defines the classification boundary. The kernel function is defined as follows:

$$K(X, Y) = \exp(-\xi^* \|X - Y\|^2) \quad (5)$$

where X and Y are the input data points, $\|\cdot\|$ represents the Euclidean distance between the data point, and ξ referred to as the kernel parameter that determines the smoothness of the decision boundary. γ is a crucial parameter for the SVM with RBF kernel, as it decides the kernel function's width, which influences the classification boundary of the data points. A lower ξ value smoothens the decision boundary of the classification task, while a higher ξ value makes the model fit the training data more closely. Besides, the larger ξ value also causes the model to overfit, in which the model fits too closely to the training data, degrading the classification performance on unseen data.

Furthermore, the regularization parameter, C, which controls the trade-off between maximizing the margin and minimizing the classification error, is another important SVM parameter that influences classification performance. In the present study, the adopted feature ranking algorithm is experimentally evaluated by the SVM classifier using ξ and C parameters of values [300, 100, 30, 10, 3, 1, 0.3, 0.1, 0.03, 0.01, 0.003], that measure the misclassification cost. For each iteration, the mentioned values are passed in sequence, yielding $11 \times 11 = 121$ combinations and validating the results with 10×5 cross-validations to determine accuracy.

2.5.2. Assessment metrics

Table 5 resumes the metrics used to assess the performance of the trained models, with a brief description and an explanation of each metric's efficiency.

3. Results and Discussion

Several tests were carried out to evaluate the proposed framework. The experimental and assessment procedures and the outcomes obtained will be provided in the subsequent sections. Furthermore, detailed explanation of the various values of deviance is also provided including details on the evaluation involved in selecting wavelet mother, decomposition level, and classifier type.

3.1. Experimentation and evaluation

In the present work, SVM is used to classify the fire pixels from the image dataset using a refined feature subset that gives higher results, as shown in Fig. 6. The 300x300 pixels 2D images are converted into time series signals for various projection angles using the Radon transform. The 37 adopted projection angles are in the range [00–3600] with a 10° step. Herein, a sampling frequency, a window size (winsize), and a window spacing (wininc) were set to 512, 128, and 512, respectively. Since the Radon transform projections, size is 429 elements; the chosen configuration allows to consider only the first window for every projection angle. It is evident that extracting the most significant coefficient is sufficient while avoiding information redundancy. Thus, the size of the final feature vector depends on the choice of the

Table 5
Assessment metrics used in this study.

Metric	Formula	Description	Advantages/ Disadvantages
Overall accuracy	$\frac{TP}{TP + TN + FP + FN}$	The number of samples accurately predicted through the dataset	It is a descriptive metric for balanced data, but it becomes deceptive for unbalanced datasets. Where FP and TP stand for false and true positives, respectively.
Specificity	$\frac{TN}{TP + TN}$	The ratio of the correctly labeled negatives over all the negatives.	It is often used when the negative accuracy rate is essential.
Precision	$\frac{TP}{TP + FP}$	The ratio of true positives over the predicted set of positives.	It is often used in cases where it is essential to have a high false positive (high cost of FPs).
Sensitivity/ Recall	$\frac{TP}{TP + FN}$	The percentage of accurately predicted true positives is also known as the true positive rate.	The model's parameters and hyperparameters directly impact the precision and recall. However, a low recall value and great precision go hand in hand.
F-measure	$\frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$	It is a mean measure of recall and precision.	Adopts a weighting policy to adjust recall and precision measures.
G-mean	$\sqrt{Precision \cdot Recall}$	The geometric mean is the product's root of the sensitivity and specificity.	This metric considers the issue of unbalanced classes.

decomposition tree level (J). For instance, for a 5-level tree ($J = 5$), five features per window from each channel are gathered five features per window from each channel. Accordingly, we obtain $6 \times 5 = 30$ features/channel since we have opted to extract 5 different kinds of features. For 37 projections, 30 features are extracted following the used configuration. The features are concatenated into a single vector, resulting in a vector of 1110 elements ($30 \times 37 = 1110$) for this case, as shown in Fig. 7. Further examples are presented in Fig. 7., namely for the cases of $J = 4, 6$, and 7 , as it will be the four cases that the corresponding experiments results are presented in this paper. The predefined angle projections set is [00:3600] with a step of 100, F1, ..., FX; X = {925, 1110, 1295, 1480} is the final feature vector that the size depends principally on the choice of the wavelet decomposition level J and FS is the set of selected features. A ranking approach then compresses the feature set for 30, 50, or 80 features (FS). Furthermore, an SVM approach is used for classification.

3.2. Feature extraction

All of the recovered feature vectors gathered from each window for each image plane of the 37 angles of the radon transform were concatenated to create the final feature vector.

The resulting concatenated feature vector is normalized for each occurrence using the z-score formula:

$$F' = \frac{F - \text{mean}(F)}{\text{std}(F)} \quad (6)$$

According to the choice of J , the size of the resulting feature vector, F , is $X = \{925, 1110, 1295, 1480\}$. The median and standard deviation of F are, respectively, represented by $\text{mean}(F)$ and $\text{std}(F)$, and F' is the vector's normalized value.

The 8036-instance matrix representing the final dataset is normalized via an identical formula.

3.3. Performance evaluation

The present study tested a variant combination of parameters to choose the optimized and more robust configuration. The SVM classifier evaluates the effectiveness of feature selection strategies using regularization and cost factors. At each iteration, the cost and regularization factors are set to one of the following values: 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, and 300. This results in 121 combinations, which are then validated using 5x5 cross-validation.

In addition, the data is split into 80 % for training and 20 % for testing for experimentation purposes. The LIBSVM support vector machine library [11] was employed under MATLAB environment to perform a multiclass SVM classification.

3.3.1. Feature selection

An initial set of experiments was conducted using Symlet 8 mother base wavelet over the whole set of data (the three mixed sets of data); the results are shown in the Table 6. The acronyms Accv, Sp, Pr, Rc, Fm, G-m correspond to Overall accuracy, specificity, precision, recall, f-measure, and g-mean, respectively. A set of 30, 50, and 80 features were used, and the following feature selection algorithms were deployed: mRMR, CIFE, JMI, CMIM, DISR, ICAP, CONDRED, and MIFS with a ν coefficient equal to 1. The first remark to advance is that the experiments with 80 features are the ones that give the best results. However, comparing the different feature ranking approaches, we conclude that ICAP and MIFS perform better. MIFS, with 80 features and the coefficient $\nu = 1$, draws an overall accuracy of 98.90 %, a specificity of 98.93 %; a precision and a recall of 98.88 % and 98.80 %, respectively; an identical f-measure and g-mean value of 98.90 %. However, ICAP with 80 features draws an overall accuracy of 99.04 %, a specificity of 99.48 %, a precision and a recall of 99.47 % and 98.61 %, respectively, an identical f-measure and g-mean value of 99.04 %.

Hence, further experiments are run using a set of 80 features, the two ranking approaches MIFS and ICAP, and different mother wavelets.

3.3.2. Decomposition level

The choice of the decomposition level is crucial since it affects the quality of the extracted features. Next, a brief analysis of the classification performance, using different levels of decomposition for three types of mother wavelets, is conducted. For the case of Coiflet 4 wavelet (please refer to Fig. 8.) we remark that with a decomposition level equal to 4 ($J = 4$) it gives relatively low results. While for $J = 5$ and $J = 6$ it gives higher results. For the case of Daubechies 6 (please refer to Fig. 9.) better results are noticed with $J=6$ and $J = 7$. Nevertheless, for Haar mother wavelet (please refer to Fig. 10.) it is sufficient to choose a $J \geq 5$.

To conclude, after thoroughly studying features at various levels, the present study opt for five levels of wavelet decomposition. It is worth noticing that the third and first levels were critical for classification. The present study aims to train a light classifier architecture that can accurately identify between large amounts of cloud smoke and disproportionately undetected forest fire regions in a larger batch of photos. Vertical, horizontal, and diagonal detail sub-bands were included at each level to calculate the corresponding features.

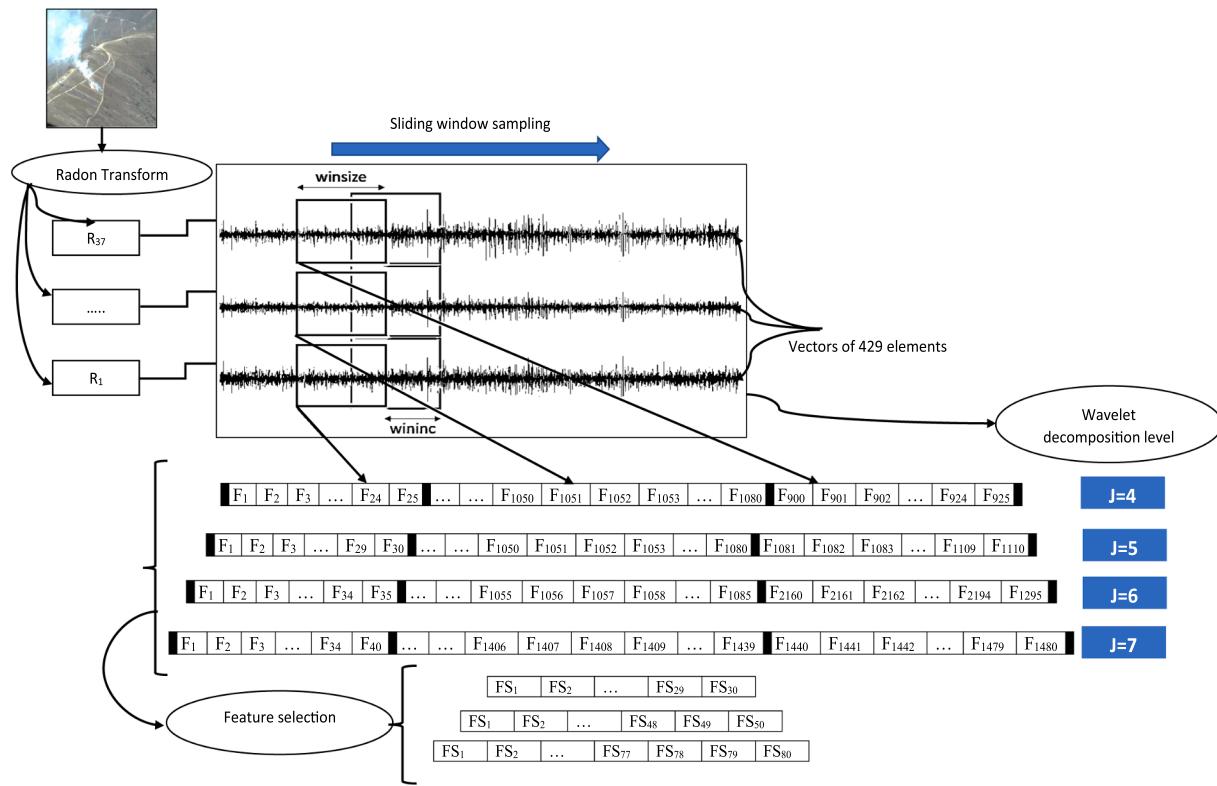


Fig. 7. Feature extraction module: *The size of every radon projection is 429 elements.*

Table 6

Classification results using Symlet 8 wavelet with 5 decomposition levels and an SVM classifier.

Feature selection technique	Number of features (i)	Overall accuracy (Acc _{ov})	Specificity (Sp)	Precision (Pr)	Recall (Rc)	F-measure (F-m)	G-mean (G-m)
mRMR	30	96.78	98.31	98.25	95.24	96.73	96.76
	50	97.93	98.98	98.96	96.89	97.91	97.93
	80	98.57	99.28	99.27	97.86	98.56	98.57
CIFE	30	94.66	96.72	96.57	92.60	94.55	94.64
	50	95.35	97.16	97.05	93.53	95.26	95.33
	80	95.89	97.81	97.72	93.97	95.81	95.87
JMI	30	94.59	95.87	95.76	93.30	94.51	94.58
	50	95.12	96.19	96.11	94.05	95.07	95.12
	80	96.49	97.41	97.36	95.57	96.46	96.49
CMIM	30	97.20	98.26	98.22	96.14	97.17	97.19
	50	97.88	98.86	98.83	96.91	97.86	97.88
	80	98.66	99.25	99.24	98.06	98.65	98.65
DISR	30	93.83	95.40	95.24	92.26	93.73	93.81
	50	95.08	96.17	96.08	94.00	95.03	95.08
	80	96.24	97.44	97.37	95.04	96.19	96.23
ICAP	30	98.31	98.83	98.82	97.78	98.30	98.31
	50	98.89	99.25	99.25	98.53	98.89	98.89
	80	99.04	99.48	99.47	98.61	99.04	99.04
CONDRED	30	87.82	89.43	89.07	86.21	87.61	87.80
	50	92.60	93.86	93.69	91.33	92.50	92.59
	80	93.21	94.75	94.58	91.66	93.10	93.19
MIFS ($\nu = 1$)	30	97.64	98.11	98.09	97.16	97.62	97.63
	50	98.52	98.56	98.55	98.48	98.52	98.52
	80	98.90	98.93	98.93	98.88	98.90	98.90

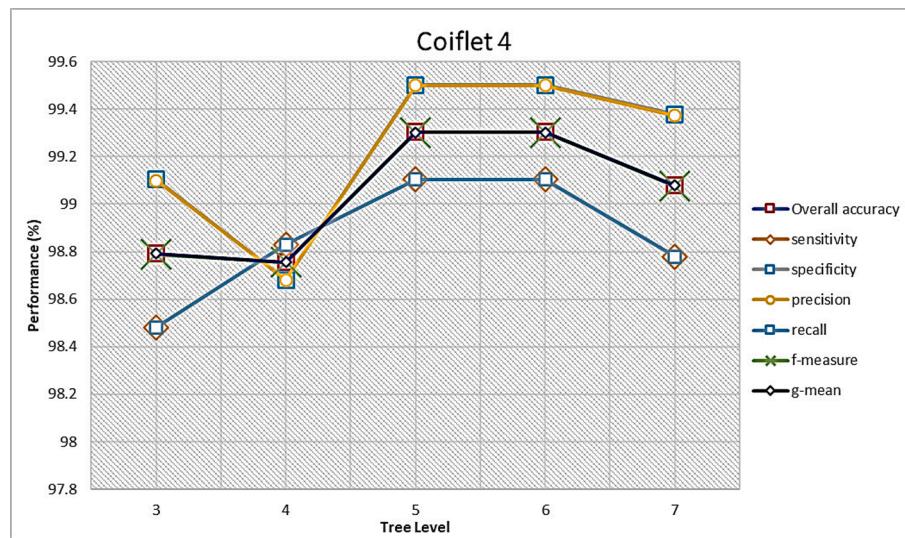


Fig. 8. Performance vs decomposition level for the case of Coiflet wavelet.

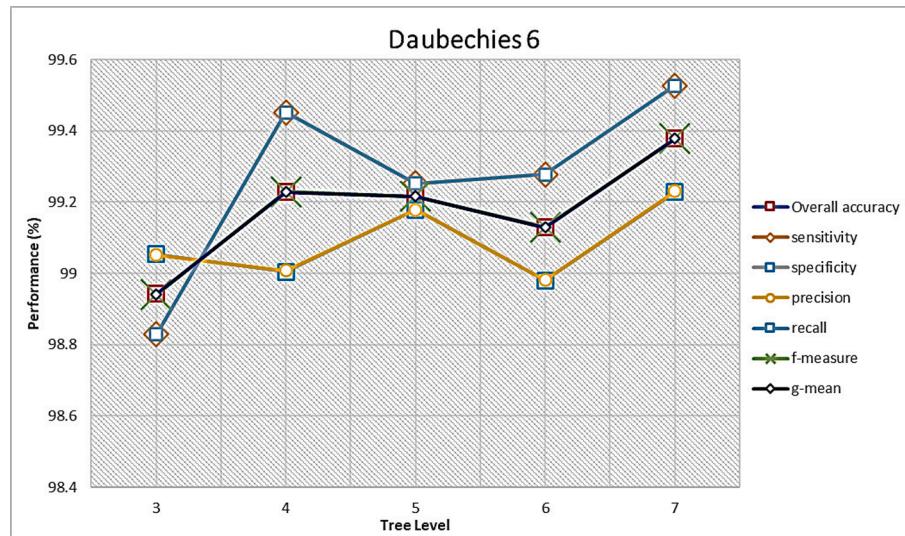


Fig. 9. Performance vs decomposition level for the case of Daubechies 6 wavelet.

3.3.3. Mother wavelet

Selecting an appropriate mother wavelet is challenging, although it is crucial for extracting a pertinent and non-redundant feature set that may be easily processed. An inappropriate selection of the mother wavelet may yield an insignificant feature set, hence degrading the classification model's performance and eventually leading to higher training costs.. For instance, the results recorded with Symlet 3, 5, and 8 wavelets with 5 decomposition levels are represented in Fig. 11.

At first glance, it is apparent that the MIFS yields superior results compared to ICAP; nonetheless, this does not constitute a viable hypothesis. The overall accuracy is not adequate as a sole metric for evaluation. Further measures that more effectively represent certain real-world scenarios and characterize the issue of unbalanced datasets should also be taken into account, namely specificity, precision, recall, f-measure, and g-mean. The main goal is to increase recall and specificity while obtaining a reasonable accuracy. Apparently, recall and precision cannot quantify the number of negatives that were mistakenly classified as positives or the number of positives that were misclassified as negatives. In practical applications, achieving a greater recall value is not necessarily translated into higher precision. Models with a significant difference among recall and precision are not well-trained or tuned. An

adjustment to recall and precision via a weighting policy presents a conundrum with the F-measure. Regarding class weighting for balanced performance, G-mean measures the level of inductive bias.

The Sym 3 mother wavelet with MIFS draws an overall accuracy of 99.12 %, a specificity of 99.05 %, a precision and a recall of 99.06 % and 99.18 %, respectively, an identical f-measure and g-mean value of 99.12 %. While the Sym 5 model with ICAP gives a lower performance in terms of overall accuracy, recall, f-measure, and g-mean. For the models trained using the features extracted deploying Sym8 mother wavelet, the model with ICAP draws higher precision and specificity values of 99.15 % with very close values of g-mean and measure (98.94 % and 98.94 % respectively for ICAP model) in comparison with the model trained with features ranked using MIFS. For the models trained using Sym 8, with the features selected with the ICAP algorithm, the results are better. The model with ICAP shows an overall accuracy of 99.04 %, a specificity of 99.48 %, a precision, and a recall of 99.47 % and 98.61 %, respectively, an identical f-measure and g-mean value of 99.04 %. The model with MIFS manifests lower accuracy, specificity, precision, g-mean, and f-measure.

Nonetheless, the results within the Daubechies wavelet support are given in Fig. 12. and the results with Haar, Coif 2, and Coif 4 are given in

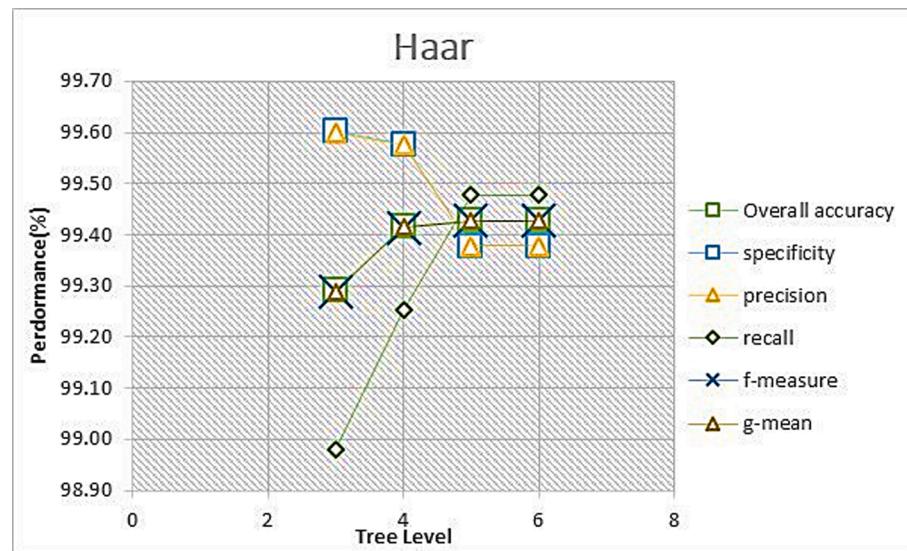


Fig. 10. Performance vs decomposition level for the case of Haar wavelet.

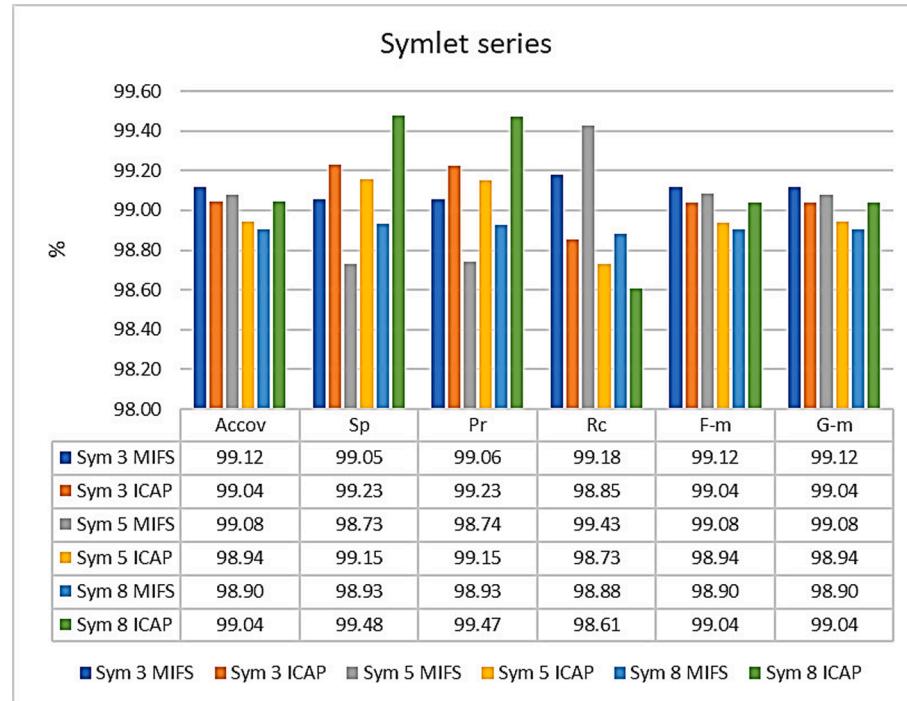


Fig. 11. Classification results using Symlet 3, 5 and 8 wavelets with 5 decomposition levels and an SVM classifier.

Fig. 13. In general, within Daubechies wavelets (Db 2, Db 4, and Db 6) MIFS algorithm performs better than ICAP. For the case of Db 2, the model with MIFS shows an overall accuracy of 99.13 %, a specificity of 98.91 %; a precision and a recall of 98.91 % and 99.35 %, respectively; an identical f-measure and g-mean value of 99.12 %. For Db 4, the model with MIFS shows similar values of overall accuracy, f-measure, and g-mean; higher specificity and precision values of 98.98 %, and a bit lower recall value of 99.28 %. However, with Db6, the recorded results show a higher value of overall accuracy of 99.22 %, a value of recall of 99.25 % (that is slightly lower than the case of Db 4), higher values of precision and specificity equal to 99.18 %, and superior values of f-measure and g-mean of 99.13 %.

It is visually evident that the higher the number of vanishing moments, the better the results.

As the same as it was noticed before, the Coif 4 demonstrates higher results than Coif 2, as shown in **Fig. 13**. Moreover, the Haar wavelet shows superior results in comparison with Coif 4 that performs better than the Symlet series. Db 6 with MIFS performs as nice as Coif 4 with ICAP with some slight differences.

A crucial characteristic is the quantity of vanishing moments in a wavelet function, which describes the wavelet's impact on diverse signals. A wavelet's number of vanishing moments, an important mathematical property, determines its capacity to suppress polynomial values. If the vanishing moment is higher, more moments (quadratic, cubic, etc.) will be extracted from the signal. The term "moments" refers to the fact that the first n derivatives of the wavelet filter's Fourier transform are all equal to zero when calculated at the sampling frequency.

This is eminently comparable to the probabilistic concept of a

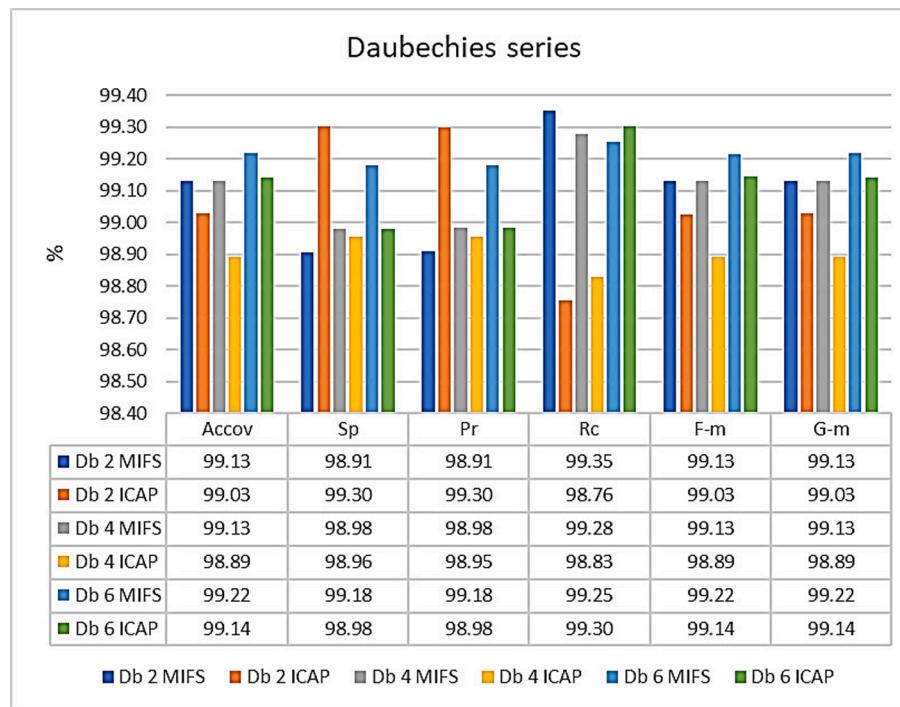


Fig. 12. The case of Daubechies 2, 4 and 6 wavelets with 5 decomposition level.

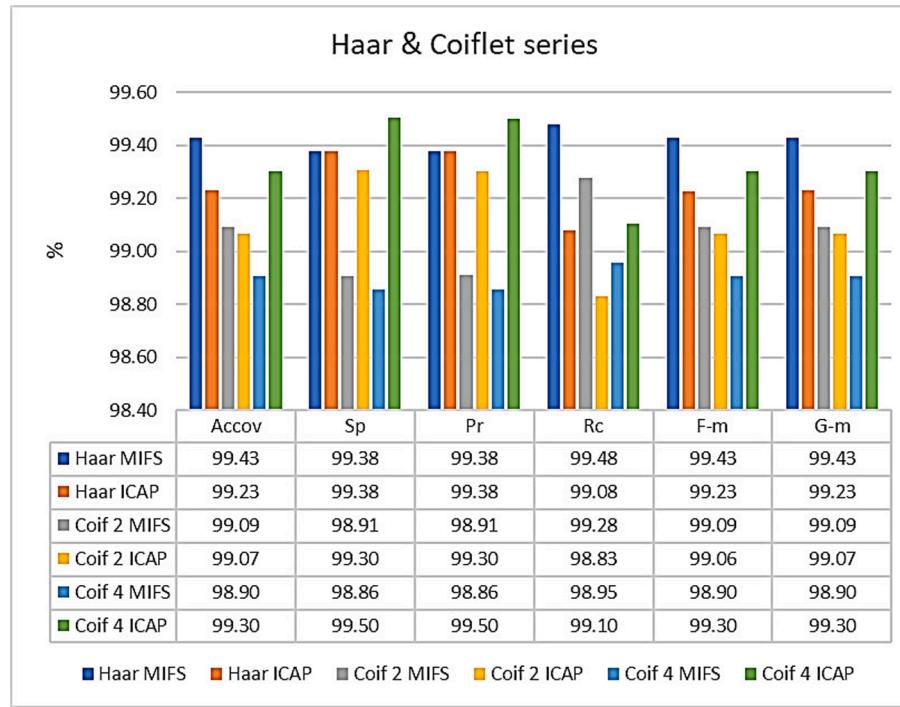


Fig. 13. The results for Haar, Coiflet 2 and 4 wavelets with 5 decomposition level.

“moment generating function” of a random variable, which is essentially the Fourier transform, and the n^{th} derivative evaluated at zero produces the n^{th} moment of the variable (i.e., the expected value, the expected value of the square, of the cube, etc.). The more the vanishing moments, the more significant and precise the collected feature set..

An example of the set of features obtained for the case of a positive (fire) and a negative (non-fire) sample is given in Fig. 14. and Fig. 15. respectively.

Inadequate mother wavelet selection may result in coefficients that do not discriminate specific features. Look at the Haar wavelet below to understand. Closely spaced features, as seen in the figures above (Fig. 14. and Fig. 15.), are effectively conveyed by detailed coefficients. Look at Sym6's to see the difference. The features are not distinguishable. With Db6 and Coif4 the features are apparent as well.

To conclude, to have a closely spaced set of features, it is mandatory to use a mother wavelet with relatively smaller support, namely Haar,

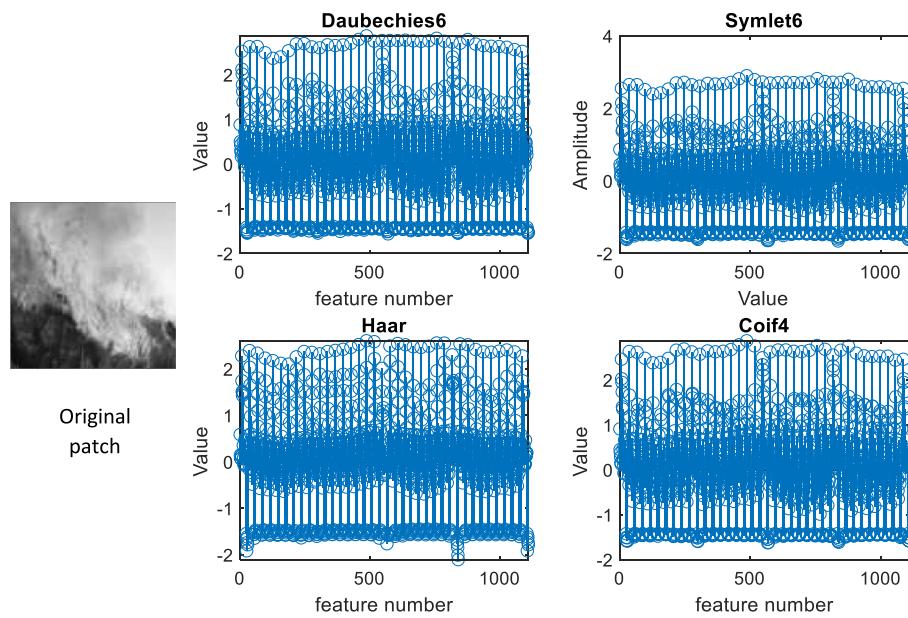


Fig. 14. Features distribution of a fire image using different type of mother wavelets: Daubechies 6, Symlet 6, Haar, and Coif 4.

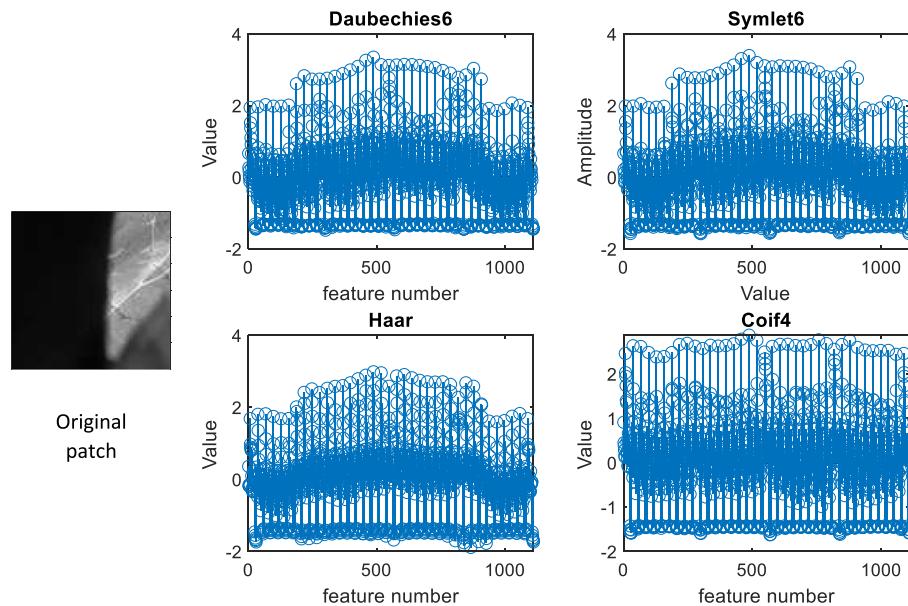


Fig. 15. Features distribution of a non-fire (background) image using different types of mother wavelets: Daubechies 6, Symlet 6, Haar, and Coif 4.

Daubechies, and Coif. In this case of figures, Haar outperforms others. In fact, smaller enough is the wavelet support more distinguished are the closely spaced feature sets.

The best models trained until now are Coif 4 with ICAP, Db 6, and

Haar with MIFS. Thus, further experiments will concentrate only on those three configurations.

Table 7

Comparison of state of the art approaches designated for early fire detection through literature.

Approach	Overall accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F-measure (%)	G-mean (%)
Detectron2 [2]	98.9	—	99.3	99.4	99.5	—
Higher Order Statistics (HOS) [22]	96.21	97.99	97.91	94.42	96.13	96.19
TuetongNet [25]	98.68	—	99.47	97.93	98.68	—
Deeplabv3 + with ResNet-50 and Dice loss [21]	98.98	—	—	91.71	94.95 %	—
YOLOv5 [34]	91.4	—	85.9	88.9	88	—
Present work (Haar with MIFS)	99.43	99.38	99.38	99.48	99.43	99.43

3.4. State of art comparison

Accurate comparison is challenging because the researchers do not reuse the same training parameters and validation processes, even though they do not use the same data, but not the same partition and assessment criteria. Table 7., highlights the most noteworthy studies and their accompanying outcomes, which could be directly compared to the present study since they use similar performance measures.

All the classified studies [2,34] draws results lower than our approach in accuracy, specificity, precision, recall, f-measure, and g-measure.

The studies deploying TuetongNet [25] and Detectron2 [2] models had pulled high results (accuracy and F-measure > 98 %) even if the specificity value of the model was not announced.

[22] introduced Higher-Order Statistics (HOS) features, which

require longer execution durations when picture dimensionality is higher. The primary distinction between the suggested method and the earlier research conducted by [22] is the intricacy of the feature extraction and refining process. The HOS feature extraction technique requires more computing power to interpolate a few erroneous values.

Using the identical set of data, we had previously studied deep learning. The model employed in the subsequently mentioned research [21] is called deeplabv3+, and it is tested with different backbones. Resnet50 backbones yield superior results. The training phase of deep learning architectures takes a long time and requires GPU devices. Here, the goal is to optimize the model's training and testing duration. Based on current research, SVM paired with a suitable collection of features is a lighter architecture that produces better results than deep learning techniques.

Nonetheless, in the Table 8. information gathered highlighting most

Table 8
Comparison of state of the art deep learning models for early fire detection through literature.

References	Feature Extraction	Dataset	Network	Accuracy
[27]	SimAM	7061 Forest fire images	Lightweight Ghost Shuffle Convolution (GSConv) with Wise Intersection over Union (WIoU) loss function.	79.6 %
[50]	Auto feature extraction using CNN and Recurrent layers (LSTM & GRU)	2947 different images obtained from Mendeley	YOLOv8, YOLOv5 and CNN-RCNN	96 % (YOLOv8 – Classification), 89 % (YOLOv8 – Object Detection) 89 % (YOLOv5-Object Detection) and 96 % (CNN-RCNN)
[57]	Auto feature extraction	DeepQuestAI(Fire, Smoke, and Neutral) dataset	Inceptionv3, MobileNetV2, and ResNet50v2	87 % (Inceptionv3), 91.33 % (MobileNetV2), 90 % (ResNet50v2)
[20]	–	1900 images (from web search)	Hybrid (ResNet152V2, InceptionV3) model and ConvNext model	99.47 % (Hybrid ResNet152V2, InceptionV3), 95.53 % (ConvNext)
SegNet approach [28]	Segmentation approach	Real-world dataset obtained by a drone flight with 7,312 images	CNN with SVM	98.2 %
Fire Segmentation-Detection Framework (PSDF) [66]	Segmentation approach	FLAME dataset and Self collected images	YOLOv8 and Vector Quantized Variational Autoencoders (VQ-VAE)	precision, recall, and F-score by 19.5 %, 1.2 %, and 11.7 % respectively. Overall accuracy of 93.5 %
[46]	Feature extracted using CNN	BoWFire with 240 images, Collected 4800 images (from MODIS, VIIRS, Copernicus Sentinel-2, and Landsat-8) and augmented to 6911 images – named as original dataset	learning without forgetting (LwF), a transfer learning approach with Xception	91.41 % – BowFire dataset, and 96.89 % – original dataset
[43]	DenseNet121 architecture	FLAME dataset (14,357 no-fire and 25,018 fire data in the training dataset, and 3,480 no-fire and 5,137 fire data in the test dataset with 1176 training, 130 validation, and 146 images were used in test sets)	InceptionV3 + GRU	99.32 %
Fire-Net [48]	Deep features using multiscale 2D convolutional layers	Landsat-8	CNN-Adam Optimizer	97.35 %
[19]	Feature Pyramidal Network	24 smoke sequences of the public HPWREN dataset	Pre-trained RetinaNet and Faster R-CNN	90 %

significant studies deploying Deep learning approaches, assessed with datasets not identical but relatively similar to the present study in terms of amount and quality of images, for performance comparison purposes. In terms of accuracy, most of the approaches [27,50,46,57,66,19,48] draws accuracy less than 98 %.

However, [20] deploys a similar performance (accuracy of 99.47 %) to our, using a transfer learning procedure and incorporating Hybrid (ResNet152V2, InceptionV3) model and ConvNexhybrid. Compared to our dataset of 8036 photos, the model was validated with a comparatively small quantity of data (1900 images that were not real scenarios).

SegNet approach [28], using CNN and SVM, draws an accuracy of 98.2 % (less than the present work's approach 99.43 %) over a dataset of 7312 set of images captured with drones.

[43] experimented DenseNet121 architecture as feature extractor and InceptionV3 + Gated recurrent unit (GRU) over the original FLAME data, distinct from the data we used for our study, which only included a subset of the dataset's scenes. The approach attends an accuracy, precision and recall of 99.32 %, 98.65 % and 100 % respectively. The technique adopted in the present research reports slightly higher recall levels, however reported lower accuracy and precision values.

However, to evaluate the model's efficiency, it is crucial to consider the model's complexity, which is calculated as the input size plus one multiplied by the number of neurons in the hidden layer, as well as the training and test times. Table 9. depicts the complexity estimation and training/test time of present work model, Deeplabv3 + with ResNet-50 and Dice loss [21], Higher Order Statistics (HOS) [22] and InceptionV3 + GRU [43].

While Deeplabv3 + demands longer training and testing durations, the present work utilized fewer computational resources. The training and testing phases of the InceptionV3 + GRU [43] took about 17 min in total. It is clear that SVM-based models are more straightforward than deep learning-based models in terms of complexity.

To our knowledge, no study has drawn similar or higher results than the current one (in terms of accuracy, specificity, precision Recall, F-measure and G-men > 99 %) over data that includes pictures with a quiet limited fire area for training and detection (aerial data). The results are highly encouraging because the fire region in the images is very limited.

4. Conclusions

Identifying and monitoring early forest fires is a difficult but necessary task. Although they are time-consuming and require specialized training, machine learning techniques—which have recently gained popularity in image processing applications and pattern recognition in the autonomous forest fire monitoring process—can effectively address the problem in this case. Therefore, conventional approaches are still valid and could eventually remedy the issue.

The present research implemented a framework based on a limited number of radon transformed features derived from adopting wavelet decomposition, notably wavelet length, standard deviation, variance, energy, and Shannon's entropy. A sliding window sampling approach is used to extract the features, which are then concatenated to generate a rather large feature vector. Ranking algorithms based on theoretical mutual information are used to select the best feature inputs for optimal model training adopting SVM with RBF kernel to reduce potential information redundancy.

The results reported based on extensive experimentation show significant performance with an increasing number of features, specifically 80 MIFS and ICAP feature inputs. Thus, further experiments were performed with the MIFS and ICAP features inputs in combination with the Symlet series (3,5,8), Daubechies series (2,4,6), Haar and Coiflet series (2,4) considering overall accuracy, specificity, precision, recall, F-measure and G-mean as the evaluation metrics. However, Haar reported better detection performance with an overall accuracy of 99.43 %, with specificity, precision, recall, F-measure, and G-mean of 99.38 %, 99.38

Table 9
Complexity/execution time comparison.

Model	Complexity Estimation			Average Execution Time	
	Number of hidden neurons/ support vectors	number of input features	Complexity	Training (Hour: Minute: Second)	Test (Second)
Deeplabv3 + with ResNet-50 and Dice loss [21]	204	512x512 pixels	53,477,580	00:55:10	0.84
Our approach (Haar with MIFS)	2080	50 features	106,080	00:02:58	0.48
Higher Order Statistics (HOS) [22]	2087	50 features	106,437	00:03:00	0.52
[43]	314 + 248	229 × 229 × 3 pixels	88,416,088	00:17:40	

%, 99.48 %, 99.43 %, and 99.43 %, respectively. Considering the restricted fire area in the aerial data set, the outcomes exhibit tremendous potential. The training time and intended model complexity are small when compared to deep learning methodologies, especially when compared to computationally expensive models like CNN.

The trained model will be tested on a new set of data acquired in a novel mission of the project to prove the efficiency of the technique further. Nonetheless, it would be interesting to consider a real-time implementation of the model over FPGA to be used in field operations.

CRediT authorship contribution statement

Houda Harkat: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hasmath Farhana Thariq Ahmed:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **José M.P. Nascimento:** Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Alexandre Bernardino:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Houda Harkat reports financial support was provided by Foundation for Science and Technology. Houda Harkat reports a relationship with Foundation for Science and Technology that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper].

Data availability

Data will be made available on request.

Acknowledgment

Houda Harkat acknowledges the financial support of the Portuguese Foundation for Science and Technology under the Project FIREFRONT with reference PCIF/SSI/0096/2017 and CTS – Center of Technology and Systems with reference UIDB/00066/2020.

Jose Nascimento acknowledges the support of the Portuguese Foundation for Science and Technology under Project UIDB/50008/2020. The authors would like to thank UAVision and the Portuguese Air Force teams that had handled the dataset acquisition task.

References

- [1] A. Abdusalomov, N. Baratov, A. Kutlimuratov, T.K. Whangbo, An improvement of the fire detection and classification method using YOLOv3 for surveillance systems, Sensors (Basel) 21 (19) (2021) 6519, <https://doi.org/10.3390/s21196519>.
- [2] A.B. Abdusalomov, B.M.S. Islam, R. Nasimov, M. Mukhiddinov, T.K. Whangbo, An improved forest fire detection method based on the detectron2 model and a deep learning approach, Sensors (Basel) 23 (3) (2023) 1512, <https://doi.org/10.3390/s23031512>.
- [3] M. Aitimov, M. Kalderova, A. Kassymova, K. Makulov, R. Muratkhan, S. Nurakynov, N. Sydyk, I. Bapiyev, Predicting and detecting fires on multispectral images using machine learning methods. Int. J. Electrical Comput. Eng. (2024) (2088-8708), 14(2).
- [4] E. Avci, I. Turkoglu, M. Poyraz, Intelligent target recognition based on wavelet packet neural network, Expert Syst. Appl. 29 (1) (2005) 175–182, <https://doi.org/10.1016/j.eswa.2005.01.016>.
- [5] C. Bahar, A. Ksibi, M. Ayadi, M.M. Jamjoom, Z. Ullah, B.O. Soufiane, H. Sakli, Wildfire and smoke detection using staged YOLO model and ensemble CNN, Electronics 12 (1) (2023) 228, <https://doi.org/10.3390/electronics12010228>.
- [6] R. Battini, Using mutual information for selecting features in supervised neural net learning, IEEE Trans. Neural Netw. 5 (4) (1994) 537–550, <https://doi.org/10.1109/72.298224>.
- [7] K. Bot, J.G. Borges, A systematic review of applications of machine learning techniques for wildfire management decision support, Inventions 7 (1) (2022) 15, <https://doi.org/10.3390/inventions7010015>.
- [8] A. Bouguettaya, H. Zarzour, A.M. Taberkit, A. Kechida, A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms, Signal Process. 190 (2022) 108309.
- [9] G. Brown, A. Pocock, M.J. Zhao, M. Luján, Conditional likelihood maximisation: a unifying framework for information theoretic feature selection, J. Mach. Learn. Res. 13 (Jan) (2012) 27–66.
- [10] Y. Cao, Q. Tang, S. Xu, F. Li, X. Lu, QuasiVSD: efficient dual-frame smoke detection, Neural Comput. & Appl. 34 (11) (2022) 8539–8550.
- [11] C.-C. Chang, C.-J. Lin, Libsvm, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27, <https://doi.org/10.1145/1961189.1961199>.
- [12] Y. Chen, W. Xu, J. Zuo, K. Yang, The fire recognition algorithm using dynamic feature fusion and IV-SVM classifier, Clust. Comput. 22 (S3) (2018) 7665–7675, <https://doi.org/10.1007/s10586-018-2368-8>.
- [13] Choudhary, C., Anurag, Shukla, P., A robust machine learning model for forest fire detection using drone images. Adv. Aerial Sens. Imaging, (2024) 129–144.
- [14] de Almeida, R. V., Crivellaro, F., Narciso, M., Sousa, A. I., & Vieira, P. (2020). Bee2Fire: A Deep Learning Powered Forest Fire Detection System. Paper presented at the ICAART (2).
- [15] T.A.T. Do, H.D. Tran, A.N.T. Do, Classifying forest cover and mapping forest fire susceptibility in Dak Nong province, Vietnam utilizing remote sensing and machine learning, Eco. Inform. 79 (2024) 102392.
- [16] F. Fleuret, Fast binary feature selection with conditional mutual information, J. Mach. Learn. Res. 5 (Nov) (2004) 1531–1555.
- [17] R. Ghosh, A. Kumar, A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire, Multimed. Tools Appl. 81 (27) (2022) 38643–38660.
- [18] J. Gubbi, S. Marusic, M. Palaniswami, Smoke detection in video using wavelets and support vector machines, Fire Saf. J. 44 (8) (2009) 1110–1115, <https://doi.org/10.1016/j.firesaf.2009.08.003>.
- [19] F. Guedé-Fernández, L. Martins, R.V. de Almeida, H. Gamboa, P. Vieira, A deep learning based object identification system for forest fire detection, Fire 4 (4) (2021) 75.
- [20] H. Gupta, N. Nihalani, An efficient fire detection system based on deep neural network for real-time applications, Signal Image and Video Processing (2024) 1–14.
- [21] H. Harkat, J. Nascimento, A. Bernardino, H. Thariq Ahmed, Assessing the impact of the loss function and encoder architecture for fire aerial images segmentation using deeplabv3+, Remote Sens. (Basel) 14 (9) (2022) 2023, <https://doi.org/10.3390/rs14092023>.
- [22] H. Harkat, J.M. Nascimento, A. Bernardino, H.F.T. Ahmed, Fire images classification based on a handcraft approach, Expert Syst. Appl. 212 (2023) 118594.
- [23] Y. Hu, J. Zhan, G. Zhou, A. Chen, W. Cai, K. Guo, L. Li, Fast forest fire smoke detection using MVMNet, Knowl.-Based Syst. 241 (2022) 108219, <https://doi.org/10.1016/j.knosys.2022.108219>.
- [24] L. Huang, G. Liu, Y. Wang, H. Yuan, T. Chen, Fire detection in video surveillances using convolutional neural networks and wavelet transform, Eng. Appl. Artif. Intel. 110 (2022) 104737, <https://doi.org/10.1016/j.engappai.2022.104737>.
- [25] G.M. Idroes, A. Maulana, R. Suhendra, A. Lala, T. Karma, F. Kusumo, T. R. Noviandy, TeutongNet: a fine-tuned deep learning model for improved forest fire detection, Leuser J. Environ. Studies 1 (1) (2023) 1–8.
- [26] A. Jakulin, Machine learning based on attribute interactions, Univerza v Ljubljani, 2005.
- [27] L. Jin, Y. Yu, J. Zhou, D. Bai, H. Lin, H. Zhou, SWVR: a lightweight deep learning algorithm for forest fire detection and recognition, Forests 15 (1) (2024) 204.
- [28] A.V. Jonnalagadda, H.A. Hashim, SegNet: a segmented deep learning based Convolutional Neural Network approach for drones wildfire detection, Remote Sens. Appl.: Soc. Environ. 34 (2024) 101181.
- [29] K.R. Mohammed, A real-time forest fire and smoke detection system using deep learning, Int. J. Nonlinear Anal. App. 13 (1) (2022) 2053–2063.
- [30] A. Khan, B. Hassan, S. Khan, R. Ahmed, A. Abuassba, H.-J. Kim, DeepFire: a novel dataset and deep transfer learning benchmark for forest fire detection, Mob. Inf. Syst. 2022 (2022) 1–14, <https://doi.org/10.1155/2022/5358359>.
- [31] J. Li, G. Zhou, A. Chen, Y. Wang, J. Jiang, Y. Hu, C. Lu, Adaptive linear feature-reuse network for rapid forest fire smoke detection model, Eco. Inform. 68 (2022) 101584, <https://doi.org/10.1016/j.ecoinf.2022.101584>.
- [32] Lin, D., & Tang, X. (2006). *Conditional infomax learning: an integrated framework for feature extraction and fusion*. Paper presented at the European Conference on Computer Vision.
- [33] J. Lin, H. Lin, F. Wang, A semi-supervised method for real-time forest fire detection algorithm based on adaptively spatial feature fusion, Forests 14 (2) (2023) 361, <https://doi.org/10.3390/f14020361>.
- [34] H. Liu, H. Hu, F. Zhou, H. Yuan, Forest flame detection in unmanned aerial vehicle imagery based on YOLOv5, Fire 6 (7) (2023) 279, <https://doi.org/10.3390/fire6070279>.
- [35] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E.S. Gündüz, K. Polat, Attention based CNN model for fire detection and localization in real-world images, Expert Syst. Appl. 189 (2022) 116114.
- [36] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693, <https://doi.org/10.1109/34.192463>.
- [37] Meyer, P. E., & Bontempi, G. (2006). *On the use of variable complementarity for feature selection in cancer classification*. Paper presented at the Workshops on applications of evolutionary computation.
- [38] Nguyen, X. V., Chan, J., Romano, S., & Bailey, J. (2014). *Effective global approaches for mutual information based feature selection*. Paper presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.
- [39] M. Park, D.Q. Tran, D. Jung, S. Park, Wildfire-detection method using DenseNet and CycleGAN data augmentation-based remote camera imagery, Remote Sens. (Basel) 12 (22) (2020) 3715.
- [40] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238, <https://doi.org/10.1109/TPAMI.2005.159>.
- [41] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert-Huang transform and wavelet transform: application to fault diagnosis for rolling bearing, Mech. Syst. Sig. Process. 19 (5) (2005) 974–988, <https://doi.org/10.1016/j.ymsp.2004.01.006>.
- [42] J. Rafiee, P.W. Tse, A. Harifi, M.H. Sadeghi, A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system, Expert Syst. Appl. 36 (3) (2009) 4862–4875, <https://doi.org/10.1016/j.eswa.2008.05.052>.
- [43] H.C. Reis, V. Turk, Detection of forest fire using deep convolutional neural networks with transfer learning approach, Appl. Soft Comput. 143 (2023) 110362.
- [44] A.C. Rowe, P.C. Abbott, Daubechies wavelets and mathematica, Comput. Phys. 9 (6) (1995) 635–648.
- [45] A. Saleh, M.A. Zukifley, H.H. Harun, F. Gaudreault, I. Davison, M. Spraggan, Forest fire surveillance systems: a review of deep learning methods, Heliyon (2024).
- [46] V.E. Sathishkumar, J. Cho, M. Subramanian, O.S. Naren, Forest fire and smoke detection using deep learning-based learning without forgetting, Fire Ecology 19 (1) (2023) 9.
- [47] Semler, L., Dettori, L., & Furst, J. (2005). *Wavelet-based texture classification of tissues in computed tomography*. Paper presented at the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05).
- [48] S.T. Seydi, V. Saeidi, B. Kalantar, N. Ueda, A.A. Halin, Fire-net: a deep learning framework for active forest fire detection, J. Sens. 2022 (1) (2022) 8044390.
- [49] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P.Z. Fulé, E. Blasch, Aerial imagery pile burn detection using deep learning: the FLAME dataset, Comput. Netw. 193 (2021) 108001, <https://doi.org/10.1016/j.comnet.2021.108001>.
- [50] I. Shama, B.E. Demir, Development of a deep learning-based surveillance system for forest fire detection and monitoring using UAV, PLoS One 19 (3) (2024) e0299058.
- [51] Singh, R., Vasquez, R. E., & Singh, R. (1997). *Comparison of Daubechies, Coiflet, and Symlet for edge detection*. Paper presented at the Visual Information Processing VI.
- [52] K. Soman, *Insight into wavelets: from theory to practice*, PHI Learning Pvt. Ltd., 2010.
- [53] Stanley, R. (1983). Deans, “The Radon Transform and Some of Its Applications.
- [54] Struzik, Z. R., & Siebes, A. (1999). *The Haar wavelet transform in the time series similarity paradigm*. Paper presented at the European Conference on Principles of Data Mining and Knowledge Discovery.

- [55] S.W. Thomas, in: *Efficient Inverse Color Map Computation*, Elsevier, 1991, pp. 116–125.
- [56] T. Toulouse, L. Rossi, A. Campana, T. Celik, M.A. Akhloufi, Computer vision for wildfire research: an evolving image dataset for processing and analysis, *Fire Saf. J.* 92 (2017) 188–194, <https://doi.org/10.1016/j.firesaf.2017.06.012>.
- [57] P. Verma, R. Bakthula, Empowering fire and smoke detection in smart monitoring through deep learning fusion, *Int. J. Inf. Technol.* 16 (1) (2024) 345–352.
- [58] J.D. Villaseñor, B. Belzer, J. Liao, Wavelet filter evaluation for image compression, *IEEE Trans. Image Process.* 4 (8) (1995) 1053–1060, <https://doi.org/10.1109/83.403412>.
- [59] T. Wang, L. Bu, Z. Yang, P. Yuan, J. Ouyang, A new fire detection method using a multi-expert system based on color dispersion, similarity and centroid motion in indoor environment, *IEEE/CAA J. Autom. Sin.* 7 (1) (2019) 263–275.
- [60] X. Wang, M. Li, M. Gao, Q. Liu, Z. Li, L. Kou, Early smoke and flame detection based on transformer, *J. Safety Sci. Resilience* 4 (3) (2023) 294–304, <https://doi.org/10.1016/j.jnlssr.2023.06.002>.
- [61] D. Wei, *Cofflet-type wavelets: theory, design, and applications*, The University of Texas at Austin, 1998.
- [62] H. Wu, D. Wu, J. Zhao, An intelligent fire detection approach through cameras based on computer vision methods, *Process Saf. Environ. Prot.* 127 (2019) 245–256, <https://doi.org/10.1016/j.psep.2019.05.016>.
- [63] X. Wu, Y. Cao, X. Lu, H. Leung, Patchwise dictionary learning for video forest fire smoke detection in wavelet domain, *Neural Comput. Applic.* 33 (13) (2021) 7965–7977, <https://doi.org/10.1007/s00521-020-05541-y>.
- [64] Yang, H., & Moody, J. (1999). *Feature selection based on joint mutual information*. Paper presented at the Proceedings of international ICSC symposium on advances in intelligent data analysis.
- [65] X. Yang, Z. Hua, L. Zhang, X. Fan, F. Zhang, Q. Ye, L. Fu, Preferred vector machine for forest fire detection, *Pattern Recogn.* 143 (2023) 109722.
- [66] H. Zhao, J. Jin, Y. Liu, Y. Guo, Y. Shen, FSDF: a high-performance fire detection framework, *Expert Syst. Appl.* 238 (2024) 121665.
- [67] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B Stat Methodol.* 67 (2) (2005) 301–320, <https://doi.org/10.1111/j.1467-9868.2005.00503.x>.