

A non-local weighted fractional order variational model for smoke detection using deep learning models

Nitish Kumar Mahala, Muzammil Khan and Pushpendra Kumar
Department of Mathematics, Bioinformatics & Computer Applications
Maulana Azad National Institute of Technology Bhopal-462003, India
223130002@stu.manit.ac.in, {mk.193104002, pkumarfma}@manit.ac.in

Abstract—As we are aware that thousands of fires break out every day around the world, which results in high numbers of casualties and serious threat to property safety and forest vegetation. Hence, it becomes particularly important to detect the fire at its early stage, because once the fire has spread in an area, it gets cataclysmic and difficult to control. In particular, the early detection of fire is associated with rising smoke. Therefore, the smoke can be considered as a good indicator to predict fire. In the presented work, smoke detection is performed with the help of its dynamical features. The dynamical features are considered in the form of optical flow color map. The motivation of this work is to use fractional order optical flow instead of images to provide the precise location and rate of growth. The estimation of optical flow is carried out using a non-local weighted fractional order variational model, which is capable in preserving dynamical discontinuities in the optical flow. Optical flow helps to find the active region of the images (video). This non-local weight also incorporates the robustness against noise. Further, the optical flow field is converted into a color map using an RGB color wheel. These color maps are used in different deep learning models for training and testing. The experiments are conducted on a dataset consisting of 18 smoke and 17 non-smoke videos. Different accuracy metrics are used for performance evaluation and detailed comparisons in order to demonstrate the significance of optical flow color maps over images in smoke detection.

Keywords Convolutional neural network, Fire prediction, Fractional order optical flow, Non-local regularization, Smoke classification.

I. INTRODUCTION

Fires are frequently occurring globally and around 300,000 casualties have been reported world wide annually, which is the fourth highest cause of accidental events [1]. Therefore, early detection of fire is very crucial in preventing extensive damage. Fire can be detected in its early stage by the utilization of smoke because in the smoldering phase of the fire, no flames are visible but smoke is produced, which is easily visible from long distances. The literature contains several traditional smoke detectors, which are based on ionization, photoelectric effect, and the presence of carbon monoxide [2]. These traditional smoke detection approaches suffer from the problems of limited detection range, utility in outdoor environments and false alarm rates [3]. However, vision based techniques offer a viable solution as these can provide a broader detection range and also feasible for outdoor fire detection. Moreover, we are aware that with the advancement in digital camera technology, surveillance cameras have become quite economic and can be easily installed in various locations such as buildings, hill stations, rural

areas, etc. These cameras store the information in the form of digital videos (image sequences). These image sequences contain many information of objects such as motion, shape and texture, etc. Therefore, when a scene contains smoke, these videos can come out as a useful tool in predicting fire breakouts. The manual processing of these substantial data volumes is an exceedingly laborious task. The literary corpus is dedicated with a significant amount of efforts in the direction of smoke detection in videos. Typically, smoke detection framework is principally classified into two approaches, which are based on: (1) traditional computer vision, (2) deep learning [4]. Traditional computer vision approaches are developed by using mathematical formulations, while the deep learning techniques approaches rely upon training and testing of a deep neural network [5], [6]. Researchers [7], [8] have presented various techniques in the literature that leverage shape, color, texture, and motion features for smoke detection, and also highlighted the challenges associated with smoke. Hanh et al. [9] conducted a study on the fire detection by utilizing aerial forest videos. They employed RGB, YCbCr, and HSI color spaces to detect the regions affected by fire. The principal issue with traditional approaches is their lack of generalization in complex scenes [10]. To tackle this problem, researchers started looking at artificial neural networks (ANNs). Since, an image contains two dimensions, which allows it to store spatial information. In order to utilize ANNs for fire detection, an image is needed to be flattened into a one dimensional vector. This dimensionality change results in the loss of crucial spatial information. In order to resolve it, works such as [11], [12] have implemented CNNs for image based fire detection. Xu et al. [11] developed a framework for smoke characterization using the deep domain adaptation techniques. Their approach leveraged synthetic data in conjunction with convolutional neural networks (CNNs) to achieve accurate smoke detection. Moreover, the fine-tuned GoogleNet architecture proposed by Muhammad et al. [12] enables the accurate detection of fire and smoke in natural videos with low computing complexity. All these approaches are based on image datasets and therefore, rely upon the static characteristics of smoke such as color, texture, shape, etc.

Nowadays, the researchers have shown that utilizing dynamic features of smoke can result in a more accurate smoke detection [13], [14]. The dynamic features of an object in a scene are generally calculated in terms of optical flow [15], [16]. Optical flow is a 2D vector plot in which

each vector corresponds to the displacement of a pixel on the object in the reference image frame. The optical flow computation is followed by an assumption that no intensity value alterations can be done on an object in an image sequence. Generally, optical flow estimation is carried out using the variational techniques [15], [16]. This is due to their simplicity in modelling a problem and accuracy in the estimated solutions [16]. Wu et al. [17] carried out with video smoke detection by implementing local binary patterns along with dense flow estimators. Their work employed HSV color conversion of flow field color maps, and specified HSV parameter values. Muller et al. [18], derived two models for optical flow computation in order to perform fire detection, which rely upon the fire saturated flames and their dynamic texture. However, the optical flow estimation models used in all these frameworks are based on integer order derivatives, which can not fully demonstrate the motion of fire-smoke due to rapid and discontinuous changes in their spatial characteristics [18]. Kumar et al. [19] generalized these variational models from integer to fractional order derivatives, and demonstrated an increased accuracy in optical flow estimation. The work such as [20] showed that smoke has a tendency of moving upwardly and this behaviour can be successfully applied in smoke detection. Therefore, this paper computes the dynamic smoke features by imbedding fractional order derivatives into the proposed variational model. Additionally, these derivatives can be employed to calculate the fractional order derivatives of order in $(0, 2)$, and demonstrate the long-term memory effect, which makes the proposed model capable of computing non-local variations in a function [21]–[23].

The literature contains several works on smoke detection such as [11], [20]. All these works rely upon the static smoke features such as color, texture, and structure for smoke detection. The novelty of the proposed work is that it utilizes both static and dynamic smoke features. As the Smoke has a tendency to move in upward direction, therefore this movement is considered as a dynamic feature of smoke, and can be computed in terms of optical flow. Moreover, in order to compute an accurate and motion preserving optical flow field, a non-local fractional-order variational model has been developed. Further, 20 different deep-learning models have been implemented for carrying out smoke detection using the computed dynamic and static smoke features.

This paper introduces a fire-smoke detection framework based on a non-local weighted fractional order optical flow and deep learning models. Optical flow color maps provide the dynamic features of smoke, and employed in devising an approach to capture the region of interest. In a color map, the colors and their intensities are in a one-to-one corresponds with flow directions and magnitudes, respectively. Due to the discontinuity and texture present in smoke, the proposed variational model is a preferable choice for estimating its optical flow, as it is capable of demonstrating information from an arbitrarily large discontinuous region. This helps in dealing the against rapid spatial variations. Finally, the estimated optical flow color maps are used for the training and testing of deep learning models. Different accuracy metrics are considered for evaluating and comparing the

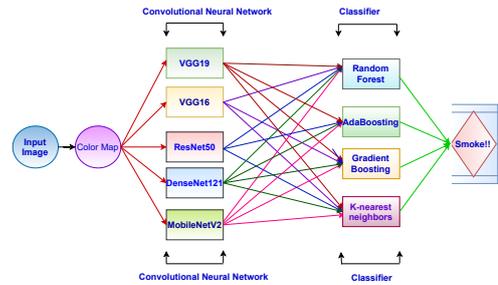


Fig. 1. Proposed framework for fire-smoke detection.

performance of the presented framework. The experiments are carried on a dataset composed of 18 smoke and 17 non-smoke videos.

II. METHODOLOGY

The overall methodology of the proposed framework is depicted in Fig. 1. The dataset used in the presented work is comprised of 18 smoke (360 frames) and 17 non-smoke (340 frames) videos, which are publicly available at [24]. Each image frame in these videos is in RGB format and of size 384×384 . These videos cover a wide range of indoor and outdoor scenes such as forests and roads with different levels of crowds, fog, brightness, and other environmental factors. Few sample reference image frames corresponding to smoke and non-smoke videos are illustrated in Fig. 2. These video frames are passed to the proposed NLWFOPF model to estimate the object motion in terms of optical flow. The optical flow is used to indicate the region of interest. Thus, a total number of 342 color maps for smoke and 323 color maps for non-smoke videos are obtained. Further, these color maps are fed into different convolutional bases and finally, classify with different machine learning classifiers.

III. NON-LOCAL WEIGHTED FRACTIONAL OPTICAL FLOW (NLWFOPF) ESTIMATION

Let $\mathcal{I}(\mathcal{X}, \mathcal{Y}, \mathcal{T})$ and $\mathcal{I}(\mathcal{X} + \delta\mathcal{X}, \mathcal{Y} + \delta\mathcal{Y}, \mathcal{T} + \delta\mathcal{T})$ be the pixel intensity values at spatiotemporal coordinates $(\mathcal{X}, \mathcal{Y}, \mathcal{T})$ and $(\mathcal{X} + \delta\mathcal{X}, \mathcal{Y} + \delta\mathcal{Y}, \mathcal{T} + \delta\mathcal{T})$, respectively. Thus, according to the brightness constancy assumption (BCA) as given in [23], we get

$$\mathcal{I}(\mathcal{X}, \mathcal{Y}, \mathcal{T}) = \mathcal{I}(\mathcal{X} + \delta\mathcal{X}, \mathcal{Y} + \delta\mathcal{Y}, \mathcal{T} + \delta\mathcal{T}) \quad (1)$$

where, $\mathcal{I} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ is a real valued function representing pixel intensity and $\Omega \subset \mathbb{R}^3$ is a volume describing the image sequence. Hence, using the Taylor series expansion in the right side of (1), we have

$$(\nabla\mathcal{I})^T \mathbf{u} + \mathcal{I}_{\mathcal{T}} = 0 \quad (2)$$

where, $\nabla\mathcal{I} = (\mathcal{I}_{\mathcal{X}}, \mathcal{I}_{\mathcal{Y}})^T$ is the spatial gradient of intensity \mathcal{I} and $\mathbf{u} = (u, v)^T$ is the optical flow with u and v as the flow components along \mathcal{X} and \mathcal{Y} directions, respectively. Also, $\mathcal{I}_{\mathcal{T}}$ denotes the temporal partial derivative of \mathcal{I} . The expression obtained in (2) is known as optical flow constraint. In order to regularize the flow field, an additional expression known as the smoothness term is added to the optical flow constraint [15]. The purpose of including this term is to reduce the effect of the outliers such as noise [25]. Thus, the proposed non-local weighted fractional order variational

model in optical flow estimation is defined as

$$E^{NM}(\mathbf{u}) = \int_{\Omega} [\mathcal{W}((\nabla \mathcal{I})^T \mathbf{u} + \mathcal{I}_{\mathcal{T}})^2 + \beta^2 H_1(u, v) + \lambda H_2(D^\alpha u, D^\alpha v)] d\mathbf{X} \quad (3)$$

where, $\mathcal{W} = \mathcal{W}(x, y)$ denotes the value of the non-local weight function at position (x, y) with respect to the center of the rectangular window. This non-local weighted function illustrates rotational invariance and can be written as

$$\mathcal{W}(x, y) = \frac{1}{\mathcal{Z}(x, y)} e^{-\frac{\|\mathcal{I}(P(x_j, y_j)) - \mathcal{I}(P(x, y))\|^2}{\sigma^2}} \quad (4)$$

Here, (x_j, y_j) denotes the j th pixel coordinate in Ω , while σ is the Gaussian variance parameter. Here, $P(x, y)$ denotes the patch value at position (x, y) , while $P(x_j, y_j)$ represents the window mask at position (x_j, y_j) , and $\mathcal{Z}(x, y)$ is a parameter for normalization and is defined as

$$\mathcal{Z}(x, y) = \sum_{(x, y) \in \Omega} e^{-\frac{\|\mathcal{I}(P(x_j, y_j)) - \mathcal{I}(P(x, y))\|^2}{\sigma^2}} \quad (5)$$

Thus, it is observed that the value of this weight is very large near to the center of the window mask, while small at the points away from the center. In the proposed variational model (3), $D^\alpha = (D_\chi^\alpha, D_{\mathcal{Y}}^\alpha)^T$ denotes the fractional derivative of order $\alpha \in (0, 1)$, and $|D^\alpha u| = \sqrt{(D_\chi^\alpha u)^2 + (D_{\mathcal{Y}}^\alpha u)^2}$ and $|D^\alpha v| = \sqrt{(D_\chi^\alpha v)^2 + (D_{\mathcal{Y}}^\alpha v)^2}$. The terms, $H_1(u, v) = |\mathbf{u}|^2$ and $H_2(D^\alpha u, D^\alpha v) = (|D^\alpha u|^2 + |D^\alpha v|^2)$ are the convex functions and known to be the flow rate driven [23]. The motivation to use fractional order derivatives instead of integer order is that these can be used even if the function is not continuous, and also more capable to preserved the texture and edge associated discontinuous information in the estimation [21], [23]. In addition to it, the fractional order models illustrate the optimal fractional order corresponding to the stable solution [23]. Thus, the proposed model can generalize the existing variational models for different values of parameters.

A. Minimization

In order to find the optical flow field $\mathbf{u} = (u, v)^T$ from the expression (3), we need to minimize the variational functional in (3) through the system of Euler-lagrange equations. Let us suppose that $u^*(\mathcal{X}, \mathcal{Y})$ and $v^*(\mathcal{X}, \mathcal{Y})$ are the required functions, and $\eta(\mathcal{X}, \mathcal{Y})$ and $\psi(\mathcal{X}, \mathcal{Y}) \in C^\infty$ are the two arbitrary functions such that

$$u(\mathcal{X}, \mathcal{Y}) = u^*(\mathcal{X}, \mathcal{Y}) + \epsilon \eta(\mathcal{X}, \mathcal{Y}) \quad v(\mathcal{X}, \mathcal{Y}) = v^*(\mathcal{X}, \mathcal{Y}) + \epsilon \psi(\mathcal{X}, \mathcal{Y})$$

Now, substituting these expressions in equation (3), we obtain

$$E^{NM}(\epsilon) = \int_{\Omega} [\mathcal{W}(\mathcal{I}_{\mathcal{X}}(u^* + \epsilon \eta) + \mathcal{I}_{\mathcal{Y}}(v^* + \epsilon \psi) + \mathcal{I}_{\mathcal{T}})^2 + \beta^2 ((u^* + \epsilon \eta)^2 + (v^* + \epsilon \psi)^2) + \lambda (|D^\alpha(u^* + \epsilon \eta)|^2 + |D^\alpha(v^* + \epsilon \psi)|^2)] d\mathbf{X} \quad (6)$$

On simplifying the expression in (6), we have

$$E^{NM}(\epsilon) = \int_{\Omega} [\mathcal{W}(\mathcal{I}_{\mathcal{X}}(u^* + \epsilon \eta) + \mathcal{I}_{\mathcal{Y}}(v^* + \epsilon \psi) + \mathcal{I}_{\mathcal{T}})^2 + \beta^2 ((u^* + \epsilon \eta)^2 + (v^* + \epsilon \psi)^2) + \lambda ((D_\chi^\alpha u^* + \epsilon D_\chi^\alpha \eta)^2 + (D_{\mathcal{Y}}^\alpha u^* + \epsilon D_{\mathcal{Y}}^\alpha \eta)^2 + (D_\chi^\alpha v^* + \epsilon D_\chi^\alpha \psi)^2 + (D_{\mathcal{Y}}^\alpha v^* + \epsilon D_{\mathcal{Y}}^\alpha \psi)^2)] d\mathbf{X} \quad (7)$$

Now, differentiating equation (7) w.r.t ϵ and putting $\epsilon = 0$, the following expression is obtained

$$E'^{NM}(0) = 2 \int_{\Omega} [\mathcal{W}(x, y)(\mathcal{I}_{\mathcal{X}} u^* + \mathcal{I}_{\mathcal{Y}} v^* + \mathcal{I}_{\mathcal{T}})(\mathcal{I}_{\mathcal{X}} \eta + \mathcal{I}_{\mathcal{Y}} \psi) + \beta^2 (u^* \eta + v^* \psi) + \lambda (D_\chi^\alpha u^* D_\chi^\alpha \eta + D_{\mathcal{Y}}^\alpha u^* D_{\mathcal{Y}}^\alpha \eta + D_\chi^\alpha v^* D_\chi^\alpha \psi + D_{\mathcal{Y}}^\alpha v^* D_{\mathcal{Y}}^\alpha \psi)] d\mathbf{X} = 0 \quad (8)$$

where, $D^{\alpha*}$ represents the right fractional derivative of order α . Thus, according to the fundamental problem of calculus of variations [21], equated to zero the coefficients of η and ψ in (8), we obtain the required system of Euler-Lagrange equations as

$$\begin{aligned} \mathcal{W}(\mathcal{I}_{\mathcal{X}} u^* + \mathcal{I}_{\mathcal{Y}} v^* + \mathcal{I}_{\mathcal{T}}) \mathcal{I}_{\mathcal{X}} + \beta^2 u^* + \lambda (D_\chi^{\alpha*} D_\chi^\alpha u^* + D_{\mathcal{Y}}^{\alpha*} D_{\mathcal{Y}}^\alpha u^*) &= 0 \quad (9) \\ \mathcal{W}(\mathcal{I}_{\mathcal{X}} u^* + \mathcal{I}_{\mathcal{Y}} v^* + \mathcal{I}_{\mathcal{T}}) \mathcal{I}_{\mathcal{Y}} + \beta^2 v^* + \lambda (D_\chi^{\alpha*} D_\chi^\alpha v^* + D_{\mathcal{Y}}^{\alpha*} D_{\mathcal{Y}}^\alpha v^*) &= 0 \quad (10) \end{aligned}$$

B. Numerical Discretization and implementation

The numerical discretization of the system of equations in (9) and (10) is carried out using the Grünwald-Letnikov (GL) definition [26]. For this purpose, let us consider the pixel position (i, j) in the image window mask of size W and the mesh grid size Δh . Thus, as per the definition of GL derivative, the fractional derivative of u can be discretized as

$$D_\chi^\alpha u(i, j) = \sum_{s=0}^{\infty} w^{(\alpha)_s} u(i-s, j) \quad (11)$$

where $w^{(\alpha)_s} = (-1)^s \binom{\alpha}{s}$ denotes the binomial expression and $w^{(\alpha)_0} = 1$, $w^{(\alpha)_s} = \left(1 - \frac{(\alpha+1)}{s}\right) w^{(\alpha)_{s-1}}$, $s = 1, 2, 3 \dots$. Thus, on using the right Riemann-Liouville definition of fractional derivative on (11), we obtain

$$D_\chi^{\alpha*} D_\chi^\alpha u(i, j) = - \sum_{s=-\infty}^0 w^{(\alpha)|s|} u(i-s, j) - \sum_{s=0}^{\infty} w^{(\alpha)_s} u(i-s, j) \quad (12)$$

Since, $\sum_{s=0}^{\infty} w^{(\alpha)_s} = 0$, therefore for the image points of view, the expression in (12) can be written as

$$D_\chi^{\alpha*} D_\chi^\alpha u(i, j) = - \sum_{s=-W}^0 w^{(\alpha)|s|} \nabla u(i-s, j) - \sum_{s=0}^W w^{(\alpha)_s} \nabla u(i-s, j) \quad (13)$$

where $\nabla u(i-s, j) = u(i-s, j) - u(i, j)$. Similarly, the derivative of u with respect to \mathcal{Y} can be represented as

$$D_{\mathcal{Y}}^{\alpha*} D_{\mathcal{Y}}^\alpha u(i, j) = - \sum_{s=-W}^0 w^{(\alpha)|s|} \nabla u(i, j-s) - \sum_{s=0}^W w^{(\alpha)_s} \nabla u(i, j-s) \quad (14)$$

Thus, the derivative discretization expression of u in combined form is given as

$$D_\chi^{\alpha*} D_\chi^\alpha u^* + D_{\mathcal{Y}}^{\alpha*} D_{\mathcal{Y}}^\alpha u^* = \sum_{(\bar{i}, \bar{j}) \in \chi(i, j)} w^{(\alpha)_{s_{\bar{i}\bar{j}}}} \{u^*(i, j) - u^*(\bar{i}, \bar{j})\} \quad (15)$$

where, χ is the set of all the pixels in the neighborhood of pixel position (i, j) in \mathcal{X} and \mathcal{Y} directions and $s_{\bar{i}\bar{j}} = \max[|\bar{i} - i|, |\bar{j} - j|]$. Similarly, the discretized expression of v is defined as

$$D_\chi^{\alpha*} D_\chi^\alpha v^* + D_{\mathcal{Y}}^{\alpha*} D_{\mathcal{Y}}^\alpha v^* = \sum_{(\bar{i}, \bar{j}) \in \chi(i, j)} w^{(\alpha)_{s_{\bar{i}\bar{j}}}} \{v^*(i, j) - v^*(\bar{i}, \bar{j})\} \quad (16)$$

Now, on simplifying the equations (9), (10), (15) and (16), the following system of equations is found

$$(\mathcal{W} \mathcal{I}_{\mathcal{X}}^2 + \beta^2 + \lambda \rho) u^* + \mathcal{W} \mathcal{I}_{\mathcal{X}} \mathcal{I}_{\mathcal{Y}} v^* = \lambda \bar{u} - \mathcal{W} \mathcal{I}_{\mathcal{X}} \mathcal{I}_{\mathcal{T}} \quad (17)$$

$$\mathcal{W}\mathcal{I}_x\mathcal{I}_y u^* + (\mathcal{W}\mathcal{I}_y^2 + \beta^2 + \lambda\rho)v^* = \lambda\bar{v} - \mathcal{W}\mathcal{I}_y\mathcal{I}_T \quad (18)$$

where,

$$\rho = \sum_{(\bar{i}, \bar{j}) \in \chi(i, j)} w^{(\alpha)_{s\bar{i}\bar{j}}}, \quad \bar{u} = \sum_{(\bar{i}, \bar{j}) \in \chi(i, j)} w^{(\alpha)_{s\bar{i}\bar{j}}} u(\bar{i}, \bar{j}) \quad \text{and}$$

$$\bar{v} = \sum_{(\bar{i}, \bar{j}) \in \chi(i, j)} w^{(\alpha)_{s\bar{i}\bar{j}}} v(\bar{i}, \bar{j})$$

C. Numerical scheme

The equations (17) and (18) are solved for obtaining the discretized system of equations corresponding to u and v as,

$$u^{(n+1)} = \frac{\lambda * \bar{u} - \mathcal{W}\mathcal{I}_{xt} - \mathcal{W}\mathcal{I}_{xy}v^{(n)}}{D_1} \quad (19)$$

$$v^{(n+1)} = \frac{\lambda * \bar{v} - \mathcal{W}\mathcal{I}_{yt} - \mathcal{W}\mathcal{I}_{xy}u^{(n+1)}}{D_2} \quad (20)$$

where $D_1 = (\mathcal{W}\mathcal{I}_x^2 + \beta^2 + \lambda\rho)$ and $D_2 = (\mathcal{W}\mathcal{I}_y^2 + \beta^2 + \lambda\rho)$. Now, the solution of these equations are determined with the help of Gauss-Seidel method [27].

IV. CONVOLUTION BASES AND ML CLASSIFIERS USED IN CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional base is a term that refers to a pretrained CNN architecture such as VGG16, VGG19, ResNet50, MobileNetV2, and DenseNet121, without their top dense layer [28]. These convolution bases are the backbone of the CNNs implemented in the presented work. Recently, these have demonstrated significant success in image recognition tasks [29]. VGG-16 and VGG-19 are two variants of VGG network, which were originally introduced by Simonyan et al. [28] with the goal of image recognition. These variants contain 16 and 19 weight layers, respectively and use the convolution filters of size 3×3 . These models can be easily implemented even in mobile devices. DenseNet121 (DNet121) architecture is amongst the smallest DenseNets, and is composed of 242 layers with 8.1 million parameters [30], while ResNet50 (RNet50) is known as a member of deep residual learning framework class. The total number of layers and parameters in it are 107 and 25.6 million, respectively. The motivation of ResNet models is that these are highly modular and help in preventing exploding and vanishing gradients [31]. MobileNetV2 is amongst the light weight neural network architecture.

In this work, for classification and comparison purposes, Ada Boosting, Gradient Boosting, K-nearest neighbor, and Random forest are employed. AdaBoost (AB) and Gradient boosting (GB) are the ensemble approaches, which utilize combinations of multiple poorly performing classifiers to increase classification accuracy [32]. K-nearest neighbours (kNN) is a lazy learner, which stores all the instances corresponding to the training samples. Also, when a test sample is encountered, the sample is analyzed with the help of k-nearest saved instances [32]. This classifier demonstrates robustness against noise. Random forest (RF) is formed of multiple random decision trees. Each tree is introduced with two types of randomness. First is implemented during the construction of a tree on the sample data, while the second is given at each node of the tree itself [32]. Hence, RF is useful for the data dimensionality reduction.

V. EXPERIMENTS, RESULTS AND DISCUSSION

A. Performance evaluation metrics

The performance of a model can be assessed in terms of different metrics provided in the literature [20]. This paper considers the accuracy rate (AR), precision (P), F1-score ($F1$), cross entropy loss (CEL) and hamming loss (HL) metrics based on the requirement of the proposed work. The thorough details on these metrics can be accessed from [20].

B. Experimental discussion

In this work, all the experiments have been carried using the MATLAB R2019a and Google Colab platform. Also, in order to perform a thorough analysis, all the results have been assessed in terms of several evaluation metrics such as accuracy rate, precision, F1-score, Hamming loss and cross entropy loss. Thus, the results are illustrated by the number of experiments.

First experiment provides the optical flow estimation results for the given dataset as shown in Fig. 2. The first and third rows of this represent the sample reference image frames, whereas the second and fourth rows manifest their respective optical flow color maps correspond to smoke and non-smoke frames. However, in performing the optical flow estimation, the values of parameters α , β and λ have been taken as 0.7, 0.0001 and 1000, respectively. Thus, despite of the presence of outliers such as illumination changes in a scene, the smoke motion is clearly visible from their corresponding color maps. Also, a sufficient discrimination is found in between smoke and non-smoke objects.

The objective of the second experiment is to show the significance of the motion features in fire-smoke detection instead of using image datasets. For this purpose, different convolutional neural networks and classifiers are employed on color maps as well as image datasets. These results are illustrated in tabular form in Tables I-X. Tables I-V describe the classification results corresponding to the color maps, while the Tables VI-X manifest the results for the image datasets. It is observed that the CNN architecture implemented with DenseNet121 and AB classifier exhibits the best results with an accuracy of 99.009% and 92.187% for color maps and raw images, respectively. Nevertheless, kNN classifier with DenseNet121 gives the better classification results corresponding to color maps and raw images with an accuracy of 93.069% and 87.500%, respectively. Also, the satisfactory results are illustrated by the GB classifier with VGG16, which are of accuracy 98.019% and 79.687% for color maps and raw images, respectively. Moreover, for RF classifier, the top accuracies are 94.059% and 73.437% for color maps and raw images, respectively, which correspond to the ResNet50. Thus, it can be seen that the AB classifier with DenseNet121 provides the overall best accuracy with color maps. This validates the significance of color map in fire-smoke detection.

The objective of third experiment is to perform the comparison of all the convolutional bases for color maps and images. Thus, in order to do this, the mean of all classifiers is taken corresponding to a particular convolutional base such as VGG16 in terms of different evaluation metrics. The estimated results are shown in Fig. 3. In this figure, the

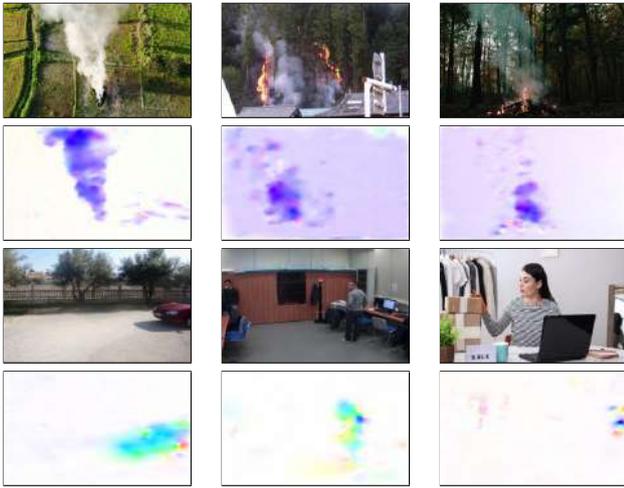


Fig. 2. Sample reference smoke and non-smoke images (first and third rows) and their corresponding optical flow color maps (second and fourth rows).

first and second bar graphs are corresponding to the optical flow color maps and raw images, respectively. These bar graphs clearly describe that on using color maps instead of images, the classification performance of a model increases significantly.

VI. CONCLUSION AND FUTURE WORK

In this work, a novel approach for the early detection of fire has introduced using the dynamical features of smoke and deep learning models. The dynamical smoke features detection has been carried out based on a non-local weighted fractional order optical flow. The obtained optical flow color maps are dense, robust, and preserve the moving edges of smoke patterns. The smoke confirmation has been performed with different combination of convolution bases and ML classifiers. A detailed comparison has been provided to differentiate in between the results of optical flow and images. It is observed that the AB classifier with DenseNet121 illustrates the overall best accuracy of 99.009% with color maps and 92.187% with images. Thus, the presented algorithm validates the proposed model. In future, the proposed algorithm can be further extended for the fog detection and weather forecasting.

TABLE I

RESULTS OF VGG19 MODEL WITH DIFFERENT CLASSIFIERS FOR COLOR MAP.

Classifiers	AR	P	F1	HL	CEL
RF	98.019	96.923	98.437	1.980	13.187
KNN	95.049	92.647	96.183	4.950	73.150
AB	98.019	96.923	98.437	1.980	37.760
GB	97.029	95.454	97.674	2.970	17.879

TABLE II

RESULTS OF VGG16 MODEL WITH DIFFERENT CLASSIFIERS FOR COLOR MAP.

Classifiers	AR	P	F1	HL	CEL
RF	97.029	95.454	97.674	2.970	13.230
KNN	94.059	91.304	95.454	5.940	41.960
AB	98.019	96.923	98.437	1.980	39.866
GB	98.019	96.923	98.437	1.980	19.629

TABLE III

RESULTS OF RESNET50 MODEL WITH DIFFERENT CLASSIFIERS FOR COLOR MAP.

Classifiers	AR	P	F1	HL	CEL
RF	94.059	92.537	95.384	5.940	24.503
KNN	91.089	87.500	93.333	8.910	78.634
AB	94.059	91.304	95.454	5.940	54.030
GB	96.039	96.825	96.825	3.960	18.985

TABLE IV

RESULTS OF DENSENET121 MODEL WITH DIFFERENT CLASSIFIERS FOR COLOR MAP.

Classifiers	AR	P	F1	HL	CEL
RF	98.019	96.923	98.437	1.980	15.055
KNN	93.069	90.000	94.736	6.930	14.426
AB	99.009	98.437	99.212	0.990	36.989
GB	96.039	94.029	96.923	3.960	8.389

TABLE V

RESULTS OF MOBILENETV2 MODEL WITH DIFFERENT CLASSIFIERS FOR COLOR MAP.

Classifiers	AR	P	F1	HL	CEL
RF	97.029	95.454	97.674	2.970	19.477
KNN	94.059	91.304	95.454	5.940	12.351
AB	96.039	95.384	96.875	3.960	47.127
GB	94.059	92.537	95.384	5.940	25.278

TABLE VI

RESULTS OF VGG19 MODEL WITH DIFFERENT CLASSIFIERS FOR IMAGES.

Classifiers	AR	P	F1	HL	CEL
RF	81.250	78.787	81.250	18.750	46.075
KNN	65.625	59.574	71.794	34.375	16.304
AB	81.250	77.142	81.818	18.750	62.637
GB	75.000	70.270	83.870	76.470	10.320

TABLE VII

RESULTS OF VGG16 MODEL WITH DIFFERENT CLASSIFIERS FOR IMAGES.

Classifiers	AR	P	F1	HL	CEL
RF	87.500	84.848	87.500	12.500	43.737
KNN	75.000	68.292	77.777	25.000	10.096
AB	84.375	78.378	85.294	15.625	61.589
GB	79.687	72.500	81.690	20.312	57.388

TABLE VIII

RESULTS OF RESNET50 MODEL WITH DIFFERENT CLASSIFIERS FOR IMAGES.

Classifiers	AR	P	F1	HL	CEL
RF	73.437	71.875	73.015	26.562	52.137
KNN	65.625	59.574	71.794	34.375	16.304
AB	81.250	75.675	82.352	18.750	66.047
GB	73.437	68.421	75.362	26.562	79.265

TABLE IX

RESULTS OF DENSENET121 MODEL WITH DIFFERENT CLASSIFIERS FOR IMAGES.

Classifiers	AR	P	F1	HL	CEL
RF	84.375	81.818	84.375	15.625	44.291
KNN	87.500	84.848	87.500	12.500	31.501
AB	92.187	93.333	91.803	7.812	57.752
GB	79.687	73.684	81.159	20.312	56.828

TABLE X

RESULTS OF MOBILENETV2 MODEL WITH DIFFERENT CLASSIFIERS FOR IMAGES.

Classifiers	AR	P	F1	HL	CEL
RF	87.500	82.857	87.878	12.500	42.165
KNN	84.375	75.609	86.111	15.625	13.849
AB	84.375	80.000	84.848	15.625	60.015
GB	85.937	82.352	86.153	14.062	47.098

VII. ACKNOWLEDGEMENT

Pushpendra Kumar and Nitish Kumar Mahala express their gratitude for the support received from NBHM, Mumbai

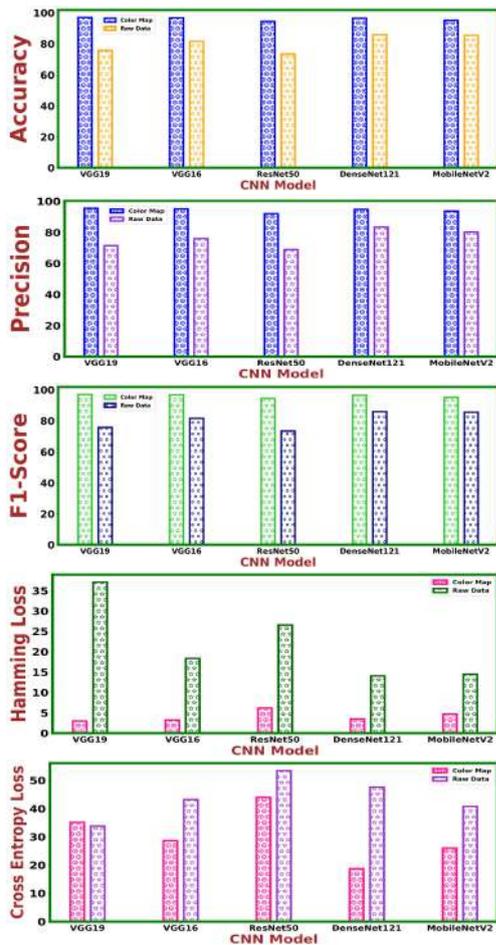


Fig. 3. Results comparison for color maps and images.

under grant no. 02011/24/2021 NBHM (R.P)/ R&D-II/8669. Also, the author Pushpendra Kumar acknowledges the support of SERB, New Delhi for grant no: EEQ/2020/000154. The author Muzammil Khan acknowledges thankfully to MHRD, New Delhi, Government of India.

REFERENCES

- [1] J. Twigg, N. Christie, J. Haworth, E. Osuteye, and A. Skarlatidou, "Improved methods for fire risk assessment in low-income and informal settlements," *International journal of environmental research and public health*, vol. 14, no. 2, p. 139, 2017.
- [2] D. T. Gottuk, M. J. Peatross, R. J. Roby, and C. L. Beyler, "Advanced fire detection using multi-signature alarm algorithms," *Fire safety journal*, vol. 37, no. 4, pp. 381–394, 2002.
- [3] A. Sharma, P. K. Singh, and Y. Kumar, "An integrated fire detection system using iot and image processing technique for smart cities," *Sustainable Cities and Society*, vol. 61, p. 102332, 2020.
- [4] G. Lin, Y. Zhang, Q. Zhang, Y. Jia, G. Xu, and J. Wang, "Smoke detection in video sequences based on dynamic texture using volume local binary patterns," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 11, no. 11, pp. 5522–5536, 2017.
- [5] T. Dawood, Z. Zhu, and T. Zayed, "Machine vision-based model for spalling detection and quantification in subway networks," *Automation in Construction*, vol. 81, pp. 149–160, 2017.
- [6] M. Mahmud, M. S. Kaiser, T. M. McGinnity, and A. Hussain, "Deep learning in mining biological data," *Cognitive computation*, vol. 13, pp. 1–33, 2021.
- [7] Y. Chunyu, F. Jun, W. Jinjun, and Z. Yongming, "Video fire smoke detection using motion and color features," *Fire technology*, vol. 46, pp. 651–663, 2010.
- [8] T. K. T. Nguyen and J.-M. Kim, "Multistage optical smoke detection approach for smoke alarm systems," *Optical Engineering*, vol. 52, no. 5, 2013.

- [9] H. Dang-Ngoc and H. Nguyen-Trung, "Aerial forest fire surveillance-evaluation of forest fire detection model using aerial videos," in *International Conference on Advanced Technologies for Communications (ATC)*, 2019, pp. 142–148.
- [10] A. Caroppo, A. Leone, and P. Siciliano, "Comparison between deep learning models and traditional machine learning approaches for facial expression recognition in ageing adults," *Journal of Computer Science and Technology*, vol. 35, pp. 1127–1146, 2020.
- [11] G. Xu, Y. Zhang, Q. Zhang, G. Lin, and J. Wang, "Deep domain adaptation based video smoke detection using synthetic smoke images," *Fire safety journal*, vol. 93, pp. 53–59, 2017.
- [12] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access*, vol. 6, pp. 18 174–18 183, 2018.
- [13] A. S. Pundir and B. Raman, "Deep belief network for smoke detection," *Fire technology*, vol. 53, pp. 1943–1960, 2017.
- [14] Y. Zhao, H. Zhang, X. Zhang, and X. Chen, "Fire smoke detection based on target-awareness and depthwise convolutions," *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 27 407–27 421, 2021.
- [15] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [16] P. Kumar and M. Khan, "Charbonnier-marchaud based fractional variational model for motion estimation in multispectral vision system," in *Journal of Physics: Conference Series*, vol. 2327, no. 1, 2022, p. 012031.
- [17] Y. Wu, M. Chen, Y. Wo, and G. Han, "Video smoke detection base on dense optical flow and convolutional neural network," *Multimedia Tools and Applications*, vol. 80, pp. 35 887–35 901, 2021.
- [18] M. Mueller, P. Karasev, I. Kolesov, and A. Tannenbaum, "Optical flow estimation for flame detection in videos," *IEEE Transactions on image processing*, vol. 22, no. 7, pp. 2786–2797, 2013.
- [19] P. Kumar, S. Kumar, and B. Raman, "A fractional order variational model for the robust estimation of optical flow from image sequences," *Optik*, vol. 127, no. 20, pp. 8710–8727, 2016.
- [20] S. Gupta, M. Khan, and P. Kumar, "Prediction of fire signatures based on fractional order optical flow and convolution neural network," in *7th International Conference on Computer Vision and Image Processing (CVIP)*, 2023, pp. 308–321.
- [21] K. Oldham and J. Spanier, *The fractional calculus theory and applications of differentiation and integration to arbitrary order*. Elsevier, 1974.
- [22] M. Khan and P. Kumar, "A nonlinear modeling of fractional order based variational model in optical flow estimation," *Optik*, vol. 261, p. 169136, 2022.
- [23] K. Muzammil and K. Pushpendra, "A level set based fractional order variational model for motion estimation in application oriented spectrum," *Expert Systems with Applications*, vol. 219, p. 119628, 2023.
- [24] *Fire Smoke dataset*. <https://www.kaggle.com/nitishkumarmahala/dataset>.
- [25] Z. Huang and A. Pan, "Non-local weighted regularization for optical flow estimation," *Optik*, vol. 208, p. 164069, 2020.
- [26] K. S. Miller and B. Ross, *An introduction to the fractional calculus and fractional differential equations*. Wiley, 1993.
- [27] R. W. Hamming, *Introduction to applied numerical analysis*. Courier Corporation, 2012.
- [28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [31] T. Garg, M. Garg, O. P. Mahela, and A. R. Garg, "Convolutional neural networks with transfer learning for recognition of covid-19: a comparative study of different approaches," *AI*, vol. 1, no. 4, pp. 586–606, 2020.
- [32] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006, vol. 4, no. 4.