

Article

Oil Spill Detection in SAR Images Using Online Extended Variational Learning of Dirichlet Process Mixtures of Gamma Distributions

Ahmed Almulihi ¹, Fahd Alharithi ¹ , Sami Bourouis ^{1,*}, Roobaea Alroobaea ¹ , Yogesh Pawar ² and Nizar Bouguila ² 

¹ College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; a.almulihi@tu.edu.sa (A.A.); f.alshalawi@tu.edu.sa (F.A.); r.robai@tu.edu.sa (R.A.)

² The Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC H3G 1T7, Canada; yogesh.pawar@mail.concordia.ca (Y.P.); nizar.bouguila@concordia.ca (N.B.)

* Correspondence: s.bourouis@tu.edu.sa

Abstract: In this paper, we propose a Dirichlet process (DP) mixture model of Gamma distributions, which is an extension of the finite Gamma mixture model to the infinite case. In particular, we propose a novel online nonparametric Bayesian analysis method based on the infinite Gamma mixture model where the determination of the number of clusters is bypassed via an infinite number of mixture components. The proposed model is learned via an online extended variational Bayesian inference approach in a flexible way where the priors of model's parameters are selected appropriately and the posteriors are approximated effectively in a closed form. The online setting has the advantage to allow data instances to be treated in a sequential manner, which is more attractive than batch learning especially when dealing with massive and streaming data. We demonstrated the performance and merits of the proposed statistical framework with a challenging real-world application namely oil spill detection in synthetic aperture radar (SAR) images.

Keywords: Dirichlet process; infinite mixture models; Gamma distribution; variational inference; online setting; oil spill detection; synthetic aperture radar images



Citation: Almulihi, A.; Alharithi, F.; Bourouis, S.; Alroobaea, R.; Pawar, Y.; Bouguila, N. Oil Spill Detection in SAR Images Using Online Extended Variational Learning of Dirichlet Process Mixtures of Gamma Distributions. *Remote Sens.* **2021**, *13*, 2991. <https://doi.org/10.3390/rs13152991>

Academic Editors: Monidipa Das, Soumya K. Ghosh, Vemuri M. Chowdary, Pabitra Mitra and Santosh Rijal

Received: 8 June 2021

Accepted: 26 July 2021

Published: 29 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of statistical machine learning has proliferated in many fields, especially to solve a broad range of problems ranging from signal processing, speech recognition, to geosciences and remote sensing where strong models are needed to apply statistical methodology. In the case of geosciences and remote sensing, for instance, statistical machine learning techniques have been deployed successfully in a variety of problems and applications in many parts of the earth system and beyond [1]. In particular, images modeling (e.g., SAR images) has received much attention due to its importance and applications in real world nature tasks related to land, climate, disturbance attribution, vegetation dynamics, urbanization, etc.

Among the probabilistic generative models, the so-named finite mixtures have been successfully applied due to their capability to represent large-scale complex probability densities and to offer a principled way for analyzing missing data [2,3]. Mixture models provide, in general, a formal approach to unsupervised learning and allow, in particular, to select the optimal number of clusters for a given dataset. This fact has been largely detailed in the literature (see, for example, [4,5]). This growing interest has led to developing several fascinating and flexible mixture models such as Gaussian-based mixture models (GMM) which have become popular even though they are not the most appropriate for fitting complex non-Gaussian shapes [6,7]. To deal with conventional GMM limitations, many other alternatives, such as Gamma (GaMM) mixtures [8–11], have shown to perform significantly better than GMM [12] thanks to its compact analytical form which is able to

cover long-tailed distributions and to approximate data with outliers. Thus, motivated by the flexibility and good performance obtained with Gamma distribution, we will focus here on investigating Gamma-based mixture model for SAR images classification. We are mainly motivated by the excellent results that Gamma mixture has provided, thanks to its flexibility, for SAR images analysis in many applications such as target detection and discrimination, target recognition and surface classification, oil spill detection, noise reduction, etc. [10]. In this paper, we will focus mainly on oil spill detection

The most challenging problem within finite mixture models is the estimation of the number of clusters that best describes the data without over- or under-fitting [13,14]. In the statistical learning context, this problem is solved using frequentist approach (i.e., maximum likelihood (ML)) within some criteria (ex. Akaike's Information Criterion, Minimum Description Length, Minimum Message Length, etc) [15,16]. It is noteworthy that the evaluation of these criteria for many clusters using ML method is very costly in terms of calculation. In addition, all parameters are supposed fixed and the inference process is based mainly on the likelihood of data which leads to convergence issues. An alternative way to tackle the issue of selecting accurately the number of clusters is via nonparametric Bayesian inference using for instance Dirichlet process (DP) [17]. In this case, the number of clusters may increase as more data are observed. This property makes DP extremely useful in exploratory data analysis. Thus, the assumption of an infinite number of components allows to avoid the problems of over- and under-fitting. Dirichlet processes (DP) mixtures have become a popular choice for various machine learning applications thanks to effective sampling techniques such as Markov chain Monte Carlo (MCMC) [18,19]. Despite the fact that MCMC yields good performance, it is frequently limited to small-scale problems and computationally intensive [20].

An interesting alternative, to both frequentist and Bayesian methods, which has provided promising performance, is variational Bayes learning [15,21]. Variational inference has the advantage to find optimal approximate posterior distributions by minimizing Kullback–Leibler (KL) divergence, or as maximizing evidence lower bound. Recently, an extended variational inference (EVI) was proposed [8] and has shown to be efficient for minimizing the KL divergence and for tackling the estimation problem. In this work, we go a step further by developing an infinite mixture model based on Gamma distribution via Dirichlet process prior, and then we propose to exploit the merits found recently by the extended variational framework [8] to learn the developed mixture model (InGaMM-eV) in an online manner. Furthermore, it is possible to estimate all parameters in closed forms. Moreover, compared to batch algorithms, online learning is more effective and helpful especially when processing big and streaming data [22] which can be crucial in SAR images analysis to allow continuous monitoring of the earth's surface. It is noteworthy also that many SAR satellite missions have accumulated repeated observations over the last decades and processing these data in an online manner could offer ease of use and solutions to some challenging problems (e.g., change detection [23]). Thus, an effective online extended variational framework of Dirichlet process mixtures of Gamma distributions is developed using stick-breaking representation. As a result, the number of clusters is selected appropriately, the model's parameters are learned in a closed form, and the issue of under-fitting is solved by deriving a model with an unlimited complexity.

The rest of this manuscript is presented as follows. We review some relevant works related to oil spill detection in Section 2. The details of extending the finite Gamma mixture to infinite case are given in Section 3. The principles of our implemented nonparametric variational learning algorithm of infinite Gamma mixture are provided in Sections 4 and 5. Section 6 is devoted to discuss the results obtained from experiments. Finally, the paper is concluded with some future works.

2. Related Research Work

Oil pollution is a major ocean disaster and environmental threat to coastal ecosystems which has been recently highlighted by several tankers accidents around the world.

Accidents on offshore oil platforms, refineries, and pipeline can cause serious oil spills. However, these accidents represent only 5% of the total oil pollution worldwide, and 95% are caused by illegal discharges by ships that prefer to dispose, cheaply, of oil residues in their tanks (according to many studies such as the European Space Agency) [24–26]. Oil pollution may result from several sources such as industrial discharges, oil production, natural oil seepage, and urban runoff. Natural slicks are of bacterial or biological decomposition or geological origin. Oil spills can devastate naval life as well as harm humans and animals by reducing dramatically air-sea exchanges processes, such as surface evaporation. Oil spills are then of great public, political and scientific concern. Therefore, there is an urgent need to monitor and detect oil spills on ocean so as to facilitate government decision making. The detection of these oil spills is considered an important and challenging problem to effectively conduct countermeasures. An effective approach is the use of satellites which provide radar images of the sea surface ($500 \times 500 \text{ km}^2$ in a single image). Satellites radar images supply an occasion to monitor coastal waters day and night, regardless of weather conditions allowing an early warning of oil spills. Moreover, satellite detection is well adapted to this kind of problems by producing images of difficult access areas [24]. Among different satellite imagery technologies, active microwave sensors such as synthetic aperture radar (SAR), has been frequently investigated for remote sensing of oil pollution [27]. The synthetic aperture radar emits and receives radio wave in order to acquire a representation of the target scene. Detecting oil spill in SAR images (as shown in Figure 1) is very complex procedure that involves many steps [26].

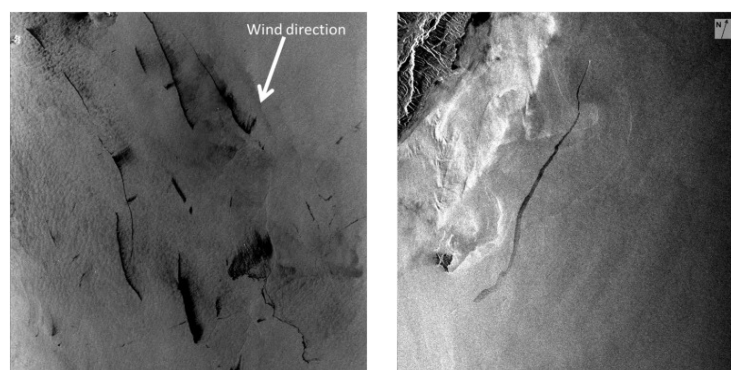


Figure 1. SAR image obtained by the European Remote Sensing satellite ERS-2 on April 1997 over the South China Sea (left image) and SAR image obtained by the ERS-1 satellite on May 1994 over Pacific Ocean east of Taiwan (right image). These images (area: $100 \text{ km} \times 100 \text{ km}$) showing an oil spill [28].

For several decades, extensive works have been provided [27,29,30] to distinguish oil slicks from natural biogenic slicks via analyzing satellite radar images. Most of conventional oil slick (or dark objects) detection procedures are carried out in three steps: (1) a pre-segmentation of dark spot, (2) the extraction of dark spot feature, and (3) a classification step of these dark spots. Some early and recent review articles summarize different oil slick detection methods [26,28,31]. These reviews state that most methods are based on using statistical patterns to discriminate between oil slicks and look-alikes under varying conditions. They conclude also that the automatic and accurate discrimination between oil spills and look-alikes is a challenging problem and need more investigations in the future. On the other side, a lot of efforts have been devoted to apply classic classifiers and descriptive statistical approaches learned from training data [25,30,32–34]. These works rely on highly trained human operators to assess and verify each region in a given SAR image. In [33], authors proposed a one-class based approach for image classification to detect oil-spill. First of all, a preprocessing step is used to identify related areas to oil spills. A feature selection step to select relevant features is also performed given that the contrast between spill's region and the surrounding regions depends on the type and amount of oil and other environmental factors (i.e., wave height, wind speed, and sea). Finally,

a one-class classifier is used to detect oil spills. A geometric level-set based segmentation method of oil spills and illegal oil discharges was developed in [35]. According to this work the regions in SAR images can be classified into pure oil spills or look-alikes on the basis of the following measurements: orientation, area, shape complexity, perimeter, eccentricity, and mean border gradient. In [36], a region-based method was also proposed. It involves both conventional detection theory and image segmentation techniques (such as N-nearest-neighbor) to have more accurate results. In [37], authors developed an adaptive thresholding-based algorithm to classify each slick as oil or look-alike. Here, involved features are derived from shape (slick complexity, width, area, moment), slick surroundings, contrast (slick local contrast, border gradient, smoothness contrast), and slick homogeneity. Their algorithms have been trained on two datasets, namely Radarsat and Envisat Advanced Synthetic Aperture Radar (ASAR) images. Fuzzy classifiers have been also used in [38] to identify all possible oil spills (dark patterns) in SAR images. A set of operations based on the fuzzy theory are used to establish the likeness of each candidate to be an oil spill or not. In the last few years, artificial neural network algorithms have been broadly applied in the context of remote sensing image segmentation and classification. Indeed, authors in [39–43] proposed different neural network-based methods (like CNN and Deep NN) in order to improve oil spill detection and classification. Some other notable interesting CNN-based oil spill detection and classification frameworks include the works in [44,45].

While considerable progress has been made in this field over the past few years, designing more robust tools still needs wide amounts of specialized knowledge and manual work. The goal here is to propose a method based on a nonparametric Bayesian model (infinite model) as well as to learn it using variational inference. Our main contributions are summarized as follow: First, we start by extending the finite Gamma mixture to the infinite case via a nonparametric Dirichlet process prior such that the problem of selecting the suitable number of clusters is solved fashionably. Then, we investigate the developed approach for remote sensing image classification. Indeed, after extracting effective features as in [46], we shall focus on modelling and classifying oil spills and other similar sea surface features using the infinite mixture model. The merits of our approach have been demonstrated using real datasets.

3. Statistical Model Specification

In this section, we present our developed variational learning approach based on the infinite Gamma mixture model.

3.1. Finite Gamma Mixture Model

Let's denote by \mathcal{Y} our observed data such as $\mathcal{Y} = \{\vec{Y}_1, \dots, \vec{Y}_N\}$, where each $\vec{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iD})$ is a D -dimensional positive vector. These feature vectors are supposed to be drawn from a mixture of Gamma distributions with parameter Θ . Let M denotes the number of mixture's components. \vec{Y}_i ($i = 1, \dots, N$) are independent and identically distributed (iid). The density function of multi-dimensional Gamma distribution is defined as follows:

$$p(\vec{Y}_i | \theta) = \prod_{d=1}^D \frac{(\beta_d)^{\alpha_d} Y_{id}^{\alpha_d-1} e^{-\beta_d Y_{id}}}{\Gamma(\alpha_d)} \quad (1)$$

where $\theta = \{\alpha_d, \beta_d\}$ is the set of parameters of the distribution such that α_d denotes the shape and β_d the location parameter. Here, $\Gamma(\cdot)$ is the Gamma function which is given as: $\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds$.

Suppose that the D -dimensional random vector \vec{Y}_i (observed data) is drawn from a finite mixture of Gamma (GaMM) distributions and consisting of M components which is

established to model the data with different shapes. The probability density function (pdf) of a GaMM is then given as:

$$p(\vec{Y} | \Theta) = p(\vec{Y} | \vec{\alpha}, \vec{\beta}, \vec{\pi}) = \prod_{i=1}^N \sum_{j=1}^M \pi_j p(\vec{Y}_i | \theta_j) \quad (2)$$

where $\Theta = \{\theta_1, \theta_2, \dots, \theta_M, \pi_1, \dots, \pi_M\}$. The parameters of the j^{th} mixture component is represented by $\theta_j = \{\alpha_j, \beta_j\}$. π_j is the vector of the mixing weights subject to $0 \leq \pi_j \leq 1$, and $\sum_{j=1}^M \pi_j = 1$.

3.2. Infinite Gamma Mixture Model

The Dirichlet process (DP) is a stochastic process with a positive scaling factor and base distribution used in Bayesian nonparametric models of data, notably in infinite mixture models. The DP is an effective concept for various applications (for more details please refer to [47]). In this section we address the issue of assuming an infinite number of components. In order to solve properly this problem which is important for well describing the observed data without over- or under-fitting, we propose a Dirichlet process mixture of Gamma distributions. In other words, we construct our infinite model by following the principle of Dirichlet process (DP) through stick-breaking representation [48,49]. Thus, the number of components is intended to be infinite. In this case, let's denote G a Dirichlet process distributed with a base distribution H and a concentration parameter ψ . The construction of $G \sim DP(\psi, H)$ is defined as

$$\begin{aligned} \lambda &\sim \text{Beta}(1, \psi) \\ \Omega_j &\sim H \\ \pi_j &= \lambda_j \prod_{s=1}^{j-1} (1 - \lambda_s) \\ G &= \sum_{j=1}^{\infty} \pi_j \delta_{\Omega_j} \end{aligned} \quad (3)$$

where δ_{Ω_j} represents the Dirac delta measure centred at Ω_j . The proportions π_j are determined by cutting a unit length stick, regularly, into an infinite number of pieces such that $\sum_{j=1}^{\infty} \pi_j = 1$ and ψ is a real number. Consequently, the infinite mixture model of Gamma distributions \mathcal{Y} is expressed as

$$p(\mathcal{Y} | \Theta) = p(\mathcal{Y} | \vec{\alpha}, \vec{\beta}, \vec{\pi}) = \prod_{i=1}^N \sum_{j=1}^{\infty} \pi_j p(\vec{Y}_i | \theta_j) \quad (4)$$

Subsequently, a latent variable $Z_i = (Z_{i1}, Z_{i2}, \dots)$ is introduced for observed data \mathcal{Y} . These latent membership vectors are used to point out if the vector \vec{Y}_i belongs to component j ($Z_{ij} = 1$) or not ($Z_{ij} = 0$). Now, the complete-data likelihood is expressed as

$$p(\mathcal{Y}, Z | \vec{\alpha}, \vec{\beta}, \vec{\pi}) = \prod_{i=1}^N \prod_{j=1}^{\infty} \pi_j^{z_{ij}} \left(p(\vec{Y}_i | \alpha_j, \beta_j) \right)^{z_{ij}} \quad (5)$$

According to the stick-breaking construction of DP (see Equation (3)), π_j can be expressed as a function of λ_j and after replacement, we have the following:

$$p(\mathcal{Z} | \vec{\lambda}) = \prod_{i=1}^N \prod_{j=1}^{\infty} \left[\lambda_j \prod_{s=1}^{j-1} (1 - \lambda_s) \right]^{z_{ij}} \quad (6)$$

The resulting complete-likelihood of the infinite Gamma mixture is finally expressed as (including latent variables):

$$p(\mathcal{Y}, Z | \vec{\alpha}, \vec{\beta}, \vec{\pi}) = \prod_{i=1}^N \prod_{j=1}^{\infty} \left[\lambda_j \prod_{s=1}^{j-1} (1 - \lambda_s) \right]^{z_{ij}} \left(p(\vec{Y}_i | \alpha_j, \beta_j) \right)^{z_{ij}} \quad (7)$$

4. Batch Variational Bayesian Learning

It is noteworthy that, when dealing with intractable models, variational inference is presented as a powerful deterministic alternative to approximate posteriors and likelihoods. In this section, we propose to develop a variational learning method to approximate inference for the DP, where the truncated stick-breaking construction [50] is applied to derive an approximate posterior and to estimate the model parameters. On the other side, we proceed by determining an approximation $Q(\Theta)$ for true posterior $p(\Theta | Y)$ such that $\Theta = \{Z, \alpha, \beta\}$. After that, we use the well-known KL divergence in order to reduce the difference between $Q(\Theta)$ and $p(\Theta | Y)$:

$$KL(Q || P) = \int Q(\Theta) \ln \left(\frac{p(\Theta | \mathcal{Y})}{Q(\Theta)} \right) d\Theta \quad (8)$$

$$KL(Q || P) = \ln(p(\mathcal{Y})) - \mathcal{L}(Q) \quad (9)$$

$$\mathcal{L}(Q) = \int Q(\Theta) \ln \left(\frac{p(\mathcal{Y}, \Theta)}{Q(\Theta)} \right) d\Theta \quad (10)$$

KL divergence attains value of zero if we have $Q(\Theta) = p(\Theta | \mathcal{Y})$ (since $KL(Q || P) \geq 0$). From Equation (9), it is possible to deduce that $\mathcal{L}(Q) \leq \ln p(\mathcal{Y})$ and so $\mathcal{L}(Q)$ is a lower bound to $\ln p(\mathcal{Y})$. However, it is difficult to solve the true posterior which cannot be directly estimated because of the complexity of calculation. We get around this matter by taking into account a restricted family of $Q(\Theta)$ that can be calculated [21]. In particular, the mean field theory [51] is adopted to factorize $Q(\Theta)$ into different tractable distributions such that $Q(\Theta) = \prod_{i=1}^M Q_i(\Theta_i)$. To maximize $\mathcal{L}(Q)$, we apply variational methodology with respect to each $Q_i(\Theta_i)$. Then, the optimal form of $Q_i(\Theta_i)$ denoted by $Q_s(\Theta_s)$ is given as

$$\ln Q_s(\Theta_s) = \langle \ln(p(\mathcal{Y}, \Theta)) \rangle_{j \neq s} + \text{const} \quad (11)$$

where $\langle \cdot \rangle_{j \neq s}$ is the expectation value of Q , with respect to all $Q_i(\Theta_i)$ excluding that case of $j = s$. It is noted that we have to take into account the truncation of the stick-breaking representation [49] to take advantage of the bound. Therefore, we take $\lambda_M = 1$ and $\pi_j = 0$ when $j > M$ which leads to $\sum_{j=1}^M \pi_j = 1$.

4.1. Prior Distributions for Parameters

To complete the probabilistic formulation, we have to place proper conjugate priors over the parameters λ , α and β . In particular, the Beta distribution is selected for the parameter λ (referring to Equation (3)) as follow

$$p(\lambda | \psi) = \prod_{j=1}^{\infty} \text{Beta}(1, \psi_j) = \prod_{j=1}^{\infty} \psi_j (1 - \lambda_j)^{\psi_j - 1} \quad (12)$$

Here, the hyperparameters of the Beta distribution is denoted by $\psi = (\psi_1, \psi_1, \dots)$ [52]. Moreover, it is possible to assign a conjugate Gamma prior to ψ :

$$p(\psi) = \mathcal{G}(\psi | a, b) = \prod_{j=1}^{\infty} \frac{b_j^{a_j}}{\Gamma(a_j)} \psi_j^{a_j - 1} e^{-b_j \psi_j} \quad (13)$$

For α and β , a prior Gamma distribution is imposed for them as suggested in [8] which is reasonable given that α and β are positives and also Gamma density is assumed to be too flexible and simple distribution to be selected as prior.

$$p(\vec{\alpha}) = \mathcal{G}(\vec{\alpha} \mid \vec{u}, \vec{v}) = \prod_{j=1}^{\infty} \prod_{d=1}^D \mathcal{G}(\alpha_{jd} \mid u_{jd}, v_{jd}) \quad (14)$$

$$p(\vec{\beta}) = \mathcal{G}(\vec{\beta} \mid \vec{s}, \vec{t}) = \prod_{j=1}^{\infty} \prod_{d=1}^D \mathcal{G}(\beta_{jd} \mid s_{jd}, t_{jd}) \quad (15)$$

Following the graphical model in Figure 1, the resulting joint distribution is expressed as

$$\begin{aligned} p(\mathcal{Y}, \Theta) &= p(\mathcal{Y}, \mathcal{Z} \mid \vec{\alpha}, \vec{\beta}) p(\mathcal{Z} \mid \vec{\lambda}) p(\vec{\lambda} \mid \vec{\psi}) p(\vec{\psi}) p(\vec{\alpha}) p(\vec{\beta}) \\ &= \prod_{i=1}^N \prod_{j=1}^{\infty} \left[\lambda_j \prod_{s=1}^{j-1} (1 - \lambda_s) \right]^{z_{ij}} \left(p(\vec{Y}_i \mid \alpha_j, \beta_j) \right)^{z_{ij}} \\ &\quad \times \prod_{j=1}^{\infty} \psi_j (1 - \lambda_j)^{\psi_j - 1} \\ &\quad \times \prod_{j=1}^{\infty} \frac{b_j^{a_j}}{\Gamma(a_j)} \psi_j^{a_j - 1} e^{-b_j \psi_j} \\ &\quad \times \prod_{j=1}^{\infty} \prod_{d=1}^D \mathcal{G}(\alpha_{jd} \mid u_{jd}, v_{jd}) \\ &\quad \times \prod_{j=1}^{\infty} \prod_{d=1}^D \mathcal{G}(\beta_{jd} \mid s_{jd}, t_{jd}) \end{aligned} \quad (16)$$

4.2. Learning Algorithm

As explained at the beginning, the objective of this work is to approximate the true posterior $p(\Theta \mid Y)$ with a new tractable approximation denoted by $Q(\Theta)$. Furthermore, the optimal solution of variational learning is reached while maximizing the lower bound w.r.t $\Theta = \{Z, \lambda, \alpha, \beta\}$. The factorization of $Q(\Theta)$ (while taking into account the truncation M) leads to following parametric form which optimal solution is presented in Appendix A:

$$Q(\Theta) = \left[\prod_{i=1}^N \prod_{j=1}^M Q(Z_{ij}) \right] \left[\prod_{j=1}^M Q(\lambda_j) Q(\psi_j) \right] \left[\prod_{j=1}^M \prod_{d=1}^D Q(\alpha_{jd}) Q(\beta_{jd}) \right] \quad (17)$$

Once the optimal variational factors are in hand, the calculation of the lower bound $\mathcal{L}(Q)$ is then straightforward. Figure 2 presents a graphical model of the proposed infinite Gamma mixture model (inGaMM). Random variables are denoted by circles and hyperparameters are represented by rounded boxes. Then, the different steps of the implemented method are summarized in Algorithm 1.

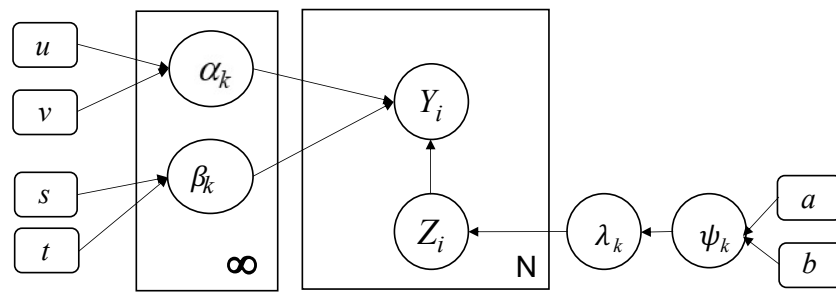


Figure 2. Graphical model of the developed variational infinite inGaMM. Random variables are denoted by circles and hyperparameters are represented by rounded boxes. Y is observed variable, Z is latent variable, large boxes are used for repeated process and the arrows show the conditional dependence between variables.

Algorithm 1: Batch variational learning approach for the inGaMM

- 1 Choose initial truncation for M .
 - 2 Set initial values for hyperparameters u, v, s, t, a, b, c, d .
 - 3 Initialize r_{ij} via k-means algorithm.
 - 4 **repeat**
 - 5 **Variational E-step :**
 - 6 Estimate the expected values according to Equations (A5), (A9), (A12), and (A15).
 - 7 **Variational M-step :**
 - 8 Update the variational solution for the factor $Q(Z)$ using Equation (A1).
 - 9 Update the variational solution for the factor $Q(\psi)$ using Equation (A6).
 - 10 Update the variational solution for the factor $Q(\lambda)$ using Equation (A7).
 - 11 Update the variational solution for the factor $Q(\alpha)$ using Equation (A10).
 - 12 Update the variational solution for the factor $Q(\beta)$ using Equation (A13).
 - 13 **until** *Until convergence is reached*
 - 14 Calculate the expected value of λ_j using Equation (A9). Then estimate the mixing coefficients according to Equation (3).
 - 15 Return the optimal number of components M_{opt} by eliminating the components with small mixing coefficients close to 0.
-

5. Online Variational Bayesian Learning

Early warning and immediate detection of oil spills has many advantages such as immediate response and reducing damage to the environment. The development of real-time monitoring and detection system is of great importance in order to minimize the volume of oil spilled. To address this problem, we propose to develop an online learning approach which is being commonly used in many other areas especially when data points are continuously arriving over time [53]. The online setting is particularly useful for incrementally training the system by feeding instances of data sequentially. It also has the benefit of making the learning process easier and faster than batch mode.

In what follows, we extend the batch variational method (presented in previous section) for unsupervised SAR images classification to an online setting. This process requires updating the model's parameters incrementally without degrading its efficiency and flexibility. To determine the lower bound, we suppose that we have at time t a fixed set of observed data. At time $t + 1$, a new SAR image Y_{N+1} comes out and is added to the dataset, hence, the mixtures' parameters have to be updated accordingly. Thus, in online setting, the lower bound at time t is expressed as in [54]:

$$\mathcal{L}^t(Q) = \frac{N}{t} \sum_{i=1}^t \int Q(\Omega) d\Omega \sum_{Z_i} \ln \left[\frac{p(\vec{Y}_i, \vec{Z}_i | \Omega)}{Q(\vec{Z}_i)} \right] + \int Q(\Omega) \ln \left[\frac{p(\Omega)}{Q(\Omega)} \right] d\Omega \quad (18)$$

where $\Omega = \{\alpha, \beta\}$.

Let's suppose that we already observed $\{\vec{Y}_1, \dots, \vec{Y}_{(t-1)}\}$ and then a new data point \vec{Y}_t is coming. Therefore, $\mathcal{L}^t(Q)$ is maximized w.r.t $Q(\vec{Z}_t)$, such that $Q(\alpha)$, $Q(\lambda)$ and $Q(\beta)$ are set to $Q^{t-1}(\alpha)$, $Q^{t-1}(\lambda)$ and $Q^{t-1}(\beta)$, respectively. We adopt a truncation technique with value M which gives [49]:

$$Q(\vec{Z}_t) = \prod_{j=1}^M r_{tj}^{Z_{tj}} \quad (19)$$

$$r_{tj} = \frac{\rho_{tj}}{\sum_{j=1}^M \rho_{tj}} \quad (20)$$

Then, $\mathcal{L}^t(Q)$ is maximized w.r.t $Q(\alpha)$, $Q(\lambda)$ and $Q(\beta)$ while keeping $Q(\vec{Z}_t)$ fixed.

$$Q^{(t)}(\vec{\alpha}) = \prod_{j=1}^M \prod_{d=1}^D \mathcal{G}(\alpha_{jd}^{(t)} | u_{jd}^{*(t)}, v_{jd}^{*(t)}) \quad (21)$$

$$Q^{(t)}(\vec{\beta}) = \prod_{j=1}^M \prod_{d=1}^D \mathcal{G}(\beta_{jd}^{(t)} | s_{jd}^{*(t)}, t_{jd}^{*(t)}) \quad (22)$$

$$Q^{(t)}(\lambda) = \prod_{j=1}^M \text{Beta}(\lambda_j^{(t)} | c_j^{(t)}, d_j^{(t)}) \quad (23)$$

where

$$\begin{aligned} u_{jd}^{*(t)} &= u_{jd}^{*(t-1)} + \rho_t \Delta u_{jd}^{*(t)} \\ v_{jd}^{*(t)} &= v_{jd}^{*(t-1)} + \rho_t \Delta v_{jd}^{*(t)} \\ s_{jd}^{*(t)} &= s_{jd}^{*(t-1)} + \rho_t \Delta s_{jd}^{*(t)} \\ t_{jd}^{*(t)} &= t_{jd}^{*(t-1)} + \rho_t \Delta t_{jd}^{*(t)} \\ c_{jd}^{*(t)} &= c_{jd}^{*(t-1)} + \rho_t \Delta c_{jd}^{*(t)} \\ d_{jd}^{*(t)} &= d_{jd}^{*(t-1)} + \rho_t \Delta d_{jd}^{*(t)} \end{aligned} \quad (24)$$

Δ is the natural gradient of each hyperparameter in the previous equation. ρ_t denotes the learning rate [55] expressed by following equation:

$$\rho_t = (\eta_0 + t)^{-\epsilon} \quad (25)$$

where $\epsilon \in [0.5, 1]$ and $\eta \geq 0$. This helps to guarantee convergence [55]. Please note that the expectation in the above mentioned equations are obtained with same manner as for the case of batch setting in the previous section and as in [56]. Since the online learning framework can be considered as a stochastic approximation algorithm, the convergence is ensured as prove in [53]. The proposed and developed online variational algorithm is presented in Algorithm 2.

Algorithm 2: Proposed online algorithm for inGaMM

```

1 Select initial truncation level  $M$ .
2 Set initial values for hyperparameters
3 Initialize  $r_{tj}$  via k-means algorithm.
4 repeat
5   Variational E-step :
6   Update the variational solution for Equation (19).
7   Variational M-step :
8   Compute the learning rate using Equation (25).
9   Calculate the hyperparameters using Equation (24).
10  Update the variational solutions for all factors  $Q^{(t)}(\alpha)$ ,  $Q^{(t)}(\beta)$  and  $Q^{(t)}(\lambda)$ 
    using Equations (21)–(23).
11  Repeat the variational E-step and V-step until new data is observed
12 until for  $t = 1$  to  $N$ 

```

6. Experimental Results**6.1. Data Sets**

The main objective of this section is to investigate our developed online extended variational learning framework of Dirichlet process mixture of Gamma distributions to detect oil spills in several SAR images. The second objective is to compare the performance of the proposed statistical framework with other methods from the state-of-art. First, it should be noted that one of the challenges is the lack of already common data sets for oil spill detection and this problem has been addressed by many relevant research communities such as [57,58]. Very limited data sets have been proposed in the literature, and therefore, it is too difficult to compare between published results since each method uses different data sets with different settings. In this work, we are essentially concerned with two challenging SAR databases. The first data set is the SAR images containing oil spills collected via the European Space Agency (ESA) database [40] which is composed of 1112 images with 5 different classes: Land, Look-alike, oil-spill, ships, and sea surface. The second one is a labelled SAR dataset taken from Sentinel-1 wave mode (TenGeoP-SARwv) [59] which includes 40,553 images with 10 different geophysical phenomena such as Pure Ocean Waves (F), Wind Streaks (G), Micro Convective Cells (H), Rain Cells (I), Biological Slicks (J), Sea Ice (K), Iceberg (L), Low Wind Area (M), Atmospheric Front (N), and Oceanic Front (O). Figures 3 and 4 show examples of images from these two datasets, respectively. For experiments, we randomly select half of the dataset as the training set and the rest for testing. In order to quantify how well SAR images are classified, we report the results in terms of average accuracy metric and false positive rate (FPR).

Modeling and classifying SAR requires powerful statistical models to represent their content (ex. color, texture). In this work we shall focus on the problem of SAR images modeling and classification via extracting local features that describe accurately input images. Indeed, feature extraction step is a part of the dimensionality reduction process that has been broadly studied in the past. It has an important role in many computer vision applications since it helps identifying the most discriminating characteristics, reducing ambiguity and enhancing the performance. However, the presence of speckle noise in synthetic aperture radar (SAR) images, as well as low-resolution between regions (surfaces) and poor contrast, make extracting relevant features too difficult. Thus, if the representative features are well extracted, then we can correctly interpret and classify images. Extracting local features from grey-scale images is a well-studied step in the fields of image processing and computer vision and various comparative measures have been studied for many years. The study of prior techniques is not within the scope of this paper. However, we suggest applying two successful methods of features extraction. The first one is based on imageNet pretrained deep learning model (resnet50) [60]. The flowchart diagram for extracting features using resnet50 is given in Figure 5. For each SAR image in the

flowchart, we first apply different image processing operations like adjusting contrast value, thresholding, object edge detection by blurring noise and small objects. After this step, based on the number of detected dark spots, we extract different features including geometrical characteristics and texture of the object. Finally, we store the extracted features for the model evaluation. In the second approach, we extract a number of features based on geometrical characteristics, physical behavior, and those related to oil spill context of the dark formations as described in [61]. After extracting features, we applied principal component analysis (PCA) to reduce dimensionality of extracted datasets features.

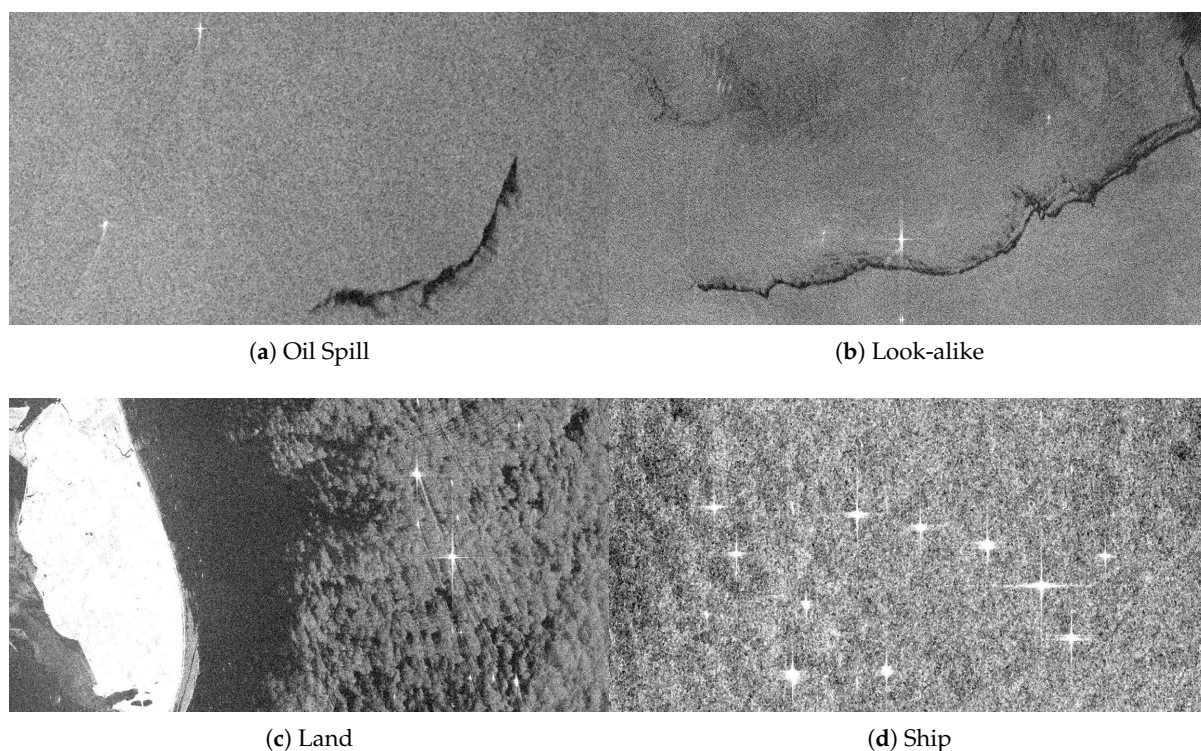
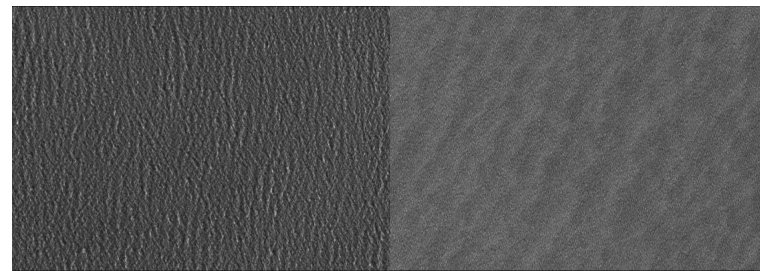


Figure 3. Dataset-1: Samples of SAR images from the European Space Agency (ESA) dataset [40]. (a) OilSpill, (b) Look-alike, (c) Land, (d) Ship.

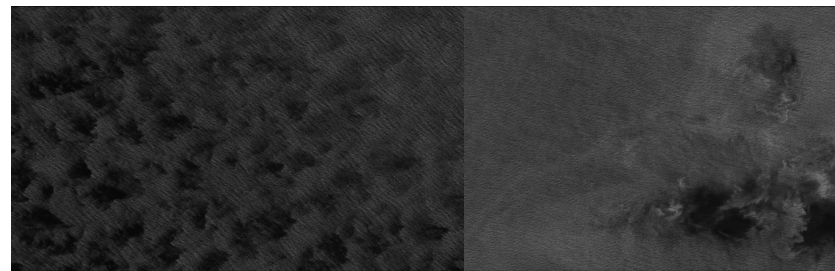
6.2. Results and Discussion

Next, we apply our online extended variational algorithm (Section 5) over the extracted features. Thus, each image is represented by an infinite Gamma mixture model. We average the results over 30 runs to evaluate and compute the final performance. Tables 1 and 2 show the average classification accuracy and false positive rate (FPR) of our InGaMM-eV model. They are obtained with different classes in both datasets and by using two features extraction methods. Indeed, we considered a first experiment where the goal was to distinguish between oil spills versus the rest and a second one where the goal is to categorize some classes from each data set (4 categories are taken from the first data set and 9 from the second one). The testing data is assumed to arrive sequentially in an online mode.



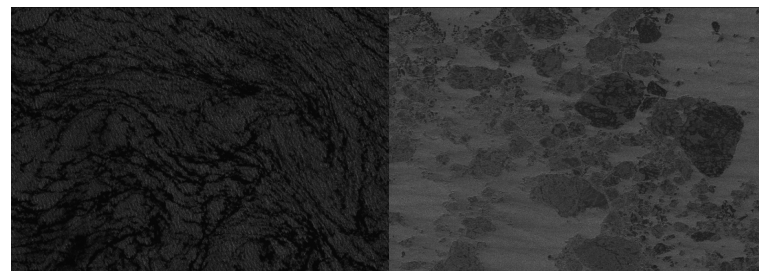
(a) Pure Ocean Waves (F)

(b) Wind Streaks (G)



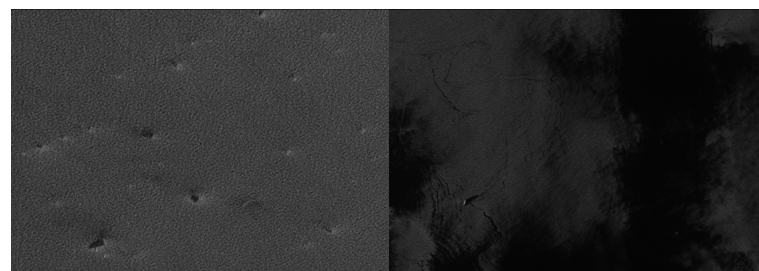
(c) Micro-Convective Cells (H)

(d) Rain Cells (I)



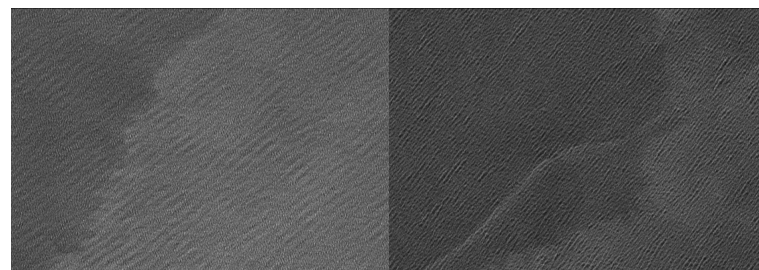
(e) Biological Slicks (J)

(f) Sea Ice (K)



(g) Iceberg (L)

(h) Low Wind Area (M)



(i) Atmospheric Front (N)

(j) Oceanic Front (O)

Figure 4. Dataset-2: Samples of SAR images from Sentinel-1 wave mode (TenGeoP-SARwv) dataset [59]. (a) Pure Ocean Waves, (b) Wind Streaks, (c) Micro-Convective Cells, (d) Rain Cells, (e) Biological Slicks, (f) Sea Ice, (g) Iceberg, (h) Low Wind Area, (i) Atmospheric Front, (j) Oceanic Front.

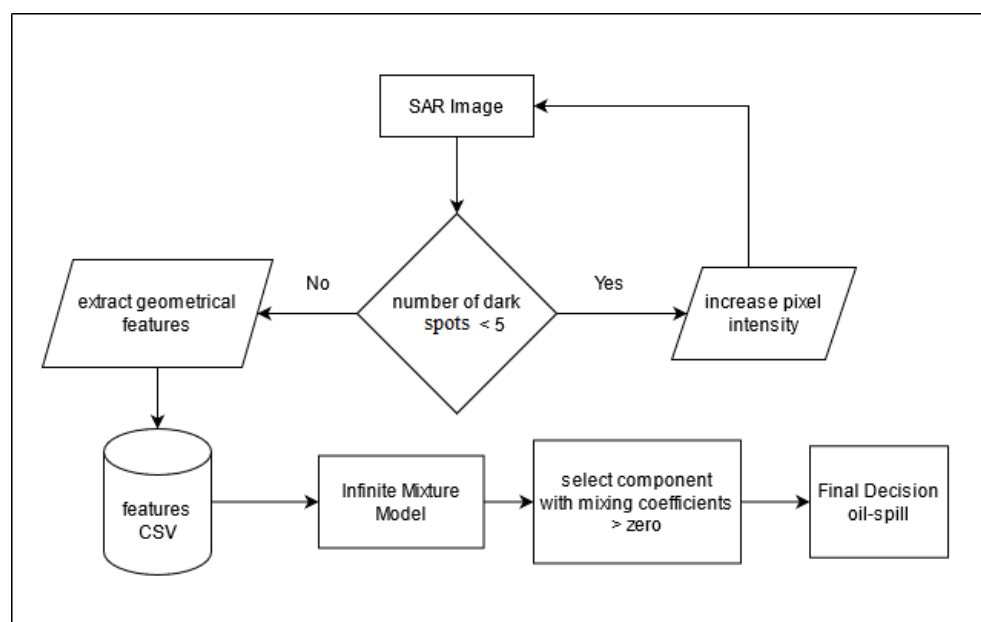


Figure 5. Flowchart diagram for extracting features using first feature extraction approach (ImageNet pretrained (resnet50) features).

Figures 6 and 7 present the confusion matrices for SAR images classification computed by the proposed InGaMM-eV using the two features extraction methods, respectively. It is noted that these matrices are used to describe the performance of the proposed model since they record true positives, false positives and false negatives. In fact, each matrix summarizes the prediction results on a classification problem and it offers a clear idea of what the proposed model is working correctly and what kinds of errors it commits. Each entry of index (u, v) represents the number of images in class u that are affected to class v . According to these results, the average classification accuracy is very promising and is equal to 90.57% (error rate of 9%) for the first dataset and 95.16% (error rate of 4%) for the second dataset.

Table 1. Results for both dataset with different number of classes using first feature extraction approach (ImageNet pretrained (resnet50) features).

Datasets	No of Class	Accuracy (%)	FPR
ESA-SAR dataset	2	97.96	0.02
ESA-SAR dataset	4	90.57	0.09
Sentinel-1 wave mode SAR dataset	2	94.53	0.05
Sentinel-1 wave mode SAR dataset	9	95.16	0.04

Table 2. Results for both dataset with different number of classes using second feature extraction approach (Dark spots, geometrical, physical, and characteristics features).

Datasets	No of Class	Accuracy (%)	FPR
ESA-SAR dataset	2	89.94	0.09
ESA-SAR dataset	4	85.13	0.12
Sentinel-1 wave mode SAR dataset	2	88.68	0.11
Sentinel-1 wave mode SAR dataset	9	82.22	0.14

Figures 6 and 7 present additional results obtained by changing the way visual features are extracted as well as the number of classes. Indeed, for the case of ESA-SAR dataset, InGaMM-eV provides high average accuracy of 97.96% using imageNet pretrained deep learning model (resnet50), and 89.94% using Dark spots, geometrical, physical char-

acteristics features. In both cases, the false positive rate is very low. For Sentinel-1 wave mode SAR dataset, the average accuracy to classify SAR images is 95.16% using resnet50, which is better than the second method for extracting features (only 88.68%). According to these results, we notice that the overall average classification accuracy is very encouraging, taking into account the complexity of treated images. It is noteworthy that, due to low resolution of images in the second dataset (Sentinel-1 wave mode), it was very difficult to extract features using the second feature extraction method (i.e., detecting dark objects). Thus, we have low accuracy than expected for this dataset.

	Ships	Oil-spill	Land	Look-alike
Ships	87	0	2	11
Oil-spill	0	98	0	2
Land	0	0	96	4
Look-alike	0	0	14	86

Figure 6. Average rounded confusion matrix (in terms of percentage) for SAR classification using InGaMM-eV for ESA-SAR dataset.

	F	G	H	I	J	L	M	N	O
F	100	0	0	0	0	0	0	0	0
G	3	97	0	0	0	0	0	0	0
H	0	0	98	2	0	0	0	0	0
I	0	0	0	88	0	10	2	0	0
J	0	0	0	0	100	0	0	0	0
L	2	0	0	0	0	100	0	0	0
M	0	0	0	0	4	0	96	0	0
N	0	0	4	14	0	0	0	82	0
O	0	0	0	0	0	0	0	0	100

Figure 7. Average rounded confusion matrix (in terms of percentage) for SAR classification using InGaMM-eV for Sentinel-1 wave mode SAR dataset.

In this experiment, our second goal is also to demonstrate the advantages of using extended variational framework over the maximum likelihood (via EM-algorithm), as well as the merits of infinite mixture model over its finite counterpart. Therefore, we compared the classification results using the following mixture models: InGaMM-eV (our infinite Gamma model using extended variational inference), GaMM-eV (finite Gamma model using extended variational learning), GaMM-EM (finite Gamma mixture model using expectation maximization learning), InGMM-eV (infinite Gaussian model using extended variational learning), and GMM-EM (finite Gaussian mixture model using expectation maximization learning). The average performances of all tested learning approaches, using the two features extraction methods, are depicted in Tables 3 and 4. We can see clearly that the extended variational approach provides better results than the EM. Furthermore, the merits of using a Dirichlet process mixtures of Gamma distributions (i.e. infinite mixture model) over a finite mixture model is clear by noting that better result was found with the infinite mixtures. In particular, in Table 3, the InGaMM-eV (90.05%) outperforms GaMM-eV (88.33%) in terms of classification accuracy rate for both datasets. On the other side, it is worth mentioning that our approach provides better results than the implemented

frameworks based on Gaussian mixtures. We can then deduce that the infinite Gamma model has better modeling and classification capability than the Gaussian when dealing with SAR images analysis.

Table 3. Overall oil spill detection rate of different models for 2 datasets using the first feature extraction approach (ImageNet pretrained (resnet50) features) .

Dataset	InGaMM-eV (<i>Our Approach</i>)	GaMM-eV	GaMM-EM	InGMM-eV	GMM-EM
ESA-SAR	90.05	88.33	86.07	83.21	83.11
Sentinel-1 wave SAR	91.12	89.40	87.02	84.14	83.99

Table 4. Overall oil spill detection rate of different models for 2 datasets using the second feature extraction approach (Dark spots, geometrical, physical, and characteristics features).

Dataset	InGaMM-eV (<i>Our Approach</i>)	GaMM-eV	GaMM-EM	InGMM-eV	GMM-EM
ESA-SAR	88.18	87.09	85.11	82.13	82.01
Sentinel-1 wave SAR	89.12	88.11	86.00	83.77	83.07

Next, The proposed learning approach (InGaMM-eV) is compared with some methods from the literature and the comparative study is presented in Table 5. As we can see, the proposed online algorithm performs better than other algorithms. Accordingly, it is important to emphasize the advantage of our developed extended variational formalism for infinite Gamma mixture, which can provide interesting results. It is also important to underline the merit of the online learning process, which is able to maintain high performance of oil spill prediction as well as handling data faster as they arrived. Moreover, it has the capacity to update the model incrementally without the need for retraining. All these results confirm that the proposed infinite Gamma mixture using the extended variational learning mode is a better choice thanks to the flexibility of the infinite Gamma mixture over the finite models. All these benefits make it more appropriate especially for SAR images classification especially in the case or large scale data sets.

Table 5. Comparative study between different methods from the literature on two datasets.

Method	Dataset	Feature Selection	Accuracy
InGaMM-eV (our approach)	ESA-SAR	ImageNet pretrained (resnet50)	97.96%
InGaMM-eV (our approach)	ESA-SAR	Dark spots, geometrical, physical features	89.94%
Fuzzy classification [62]	ESA-SAR	Georeference, Land masking, and Filtering	88%
InGaMM-eV (our approach)	Sentinel-1 SAR	ImageNet pretrained (resnet50)	94.53%
InGaMM-eV (our approach)	Sentinel-1 SAR	Dark spots, geometrical, physical features	88.68%
Convolutional neural network	Sentinel-1 SAR	Inception v3 CNN	93%
Artificial neural network [34]	Sentinel-1 SAR	Dark spot, shape features	87%
Method in [63]	Sentinel-1 SAR	Dark spot features	81%
Method in [64]	Sentinel-1 SAR	Dark spot, shape features	82.61%

7. Conclusions

In this paper an effective online nonparametric Bayesian analysis method based on Dirichlet process mixture of Gamma distributions (i.e., infinite Gamma mixture model) is developed to deal with the challenging problem of oil spill detection in SAR images. The Gamma distribution is considered because of its flexibility for semi-bounded data modelling. This framework is learned using an extended version of conventional variational inference in a flexible way which has certain advantages such as approximating the posteriors effectively in a closed form, easy assessment of convergence and easy optimization by offering a trade-off between frequentist techniques and MCMC-based ones.

An important property of our approach is that it does not need the specification of the number of mixture components in advance. The proposed online algorithm has also the benefit to allow data instances to be treated in a sequential manner, which is more attractive than batch learning especially when dealing with massive and streaming data. Through the challenging application of oil spill detection in SAR images, we have demonstrated the performance of our statistical framework, which is able to provide very encouraging results in terms of SAR images modeling and classification capabilities. As future work, we plan to integrate a feature selection mechanism into the proposed framework in order to improve more the classification accuracy. It is our hope that many other real-world applications related to image processing and machine learning can be addressed via our developed framework.

Author Contributions: Conceptualization, A.A. and S.B.; methodology, N.B. and F.A.; software, Y.P.; validation, R.A., S.B. and A.A.; formal analysis, F.A.; investigation, A.A.; resources, Y.P.; data curation, S.B.; writing—original draft preparation, A.A. and S.B.; writing—review and editing, N.B. and F.A.; visualization, R.A.; supervision, N.B.; project administration, A.A.; funding acquisition, A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deanship of Scientific Research, Taif University, Kingdom of Saudi Arabia, grant number 1-441-137.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository: The first data set: European Space Agency (ESA) database [40]. The second data set: Sentinel-1 wave mode (TenGeoP-SARvv) [59].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

(1) Optimal solution to $Q(Z)$.

$$Q(Z) = \prod_{i=1}^N \prod_{j=1}^M r_{ij}^{Z_{ij}} \quad (\text{A1})$$

where the responsibility r_{ij} can be calculated as:

$$r_{ij} = \frac{\rho_{ij}}{\sum_{j=1}^M \rho_{ij}} \quad (\text{A2})$$

such that:

$$\ln(\rho_{ij}) = \ln(\pi_j) + \sum_{d=1}^D [\mathcal{P}_{jd} + (\langle \alpha_{jd} \rangle - 1) \ln(y_{jd}) - \langle \alpha_{jd} \rangle \langle \beta_{jd} \rangle y_{jd}] \quad (\text{A3})$$

and

$$\mathcal{P}_{jd} = \alpha_{jd}^* \ln(\alpha_{jd}^*) - \alpha_{jd}^* - \ln(\alpha_{jd}^*) - \ln(\Gamma(\alpha_{jd}^*)) + \langle \ln(\alpha_{jd}) \rangle + \langle \alpha_{jd} \rangle + \langle \alpha_{jd} \rangle \langle \ln(\beta_{jd}) \rangle \quad (\text{A4})$$

where $\langle . \rangle$ refers to an expectation w.r.t. the corresponding factor and α_{jd}^* is any feasible point.

The expectation of Z_{ij} is determined as:

$$\langle Z_{ij} \rangle = r_{ij} \quad (\text{A5})$$

(2) Optimal solution to $Q(\psi)$ and $Q(\lambda)$.

$$Q(\psi) = \prod_{j=1}^M \mathcal{G}(\psi_j | a_j, b_j) \quad (\text{A6})$$

$$Q(\lambda) = \prod_{j=1}^M \text{Beta}(\lambda_j | c_j, d_j) \quad (\text{A7})$$

$$\begin{aligned} a_j^* &= a_j + 1 \\ b_j^* &= b_j - \langle \ln(1 - \lambda_j) \rangle \\ c_j^* &= 1 + \sum_{i=1}^N \langle Z_{ij} \rangle \\ d_j^* &= \langle \psi_j \rangle + \sum_{i=1}^N \sum_{s=j+1}^M \langle Z_{is} \rangle \end{aligned} \quad (\text{A8})$$

From the previous equations, we obtain the following expectations:

$$\begin{aligned} \langle \ln(\lambda_j) \rangle &= \Psi(c_j^*) - \Psi(c_j^* + d_j^*) \\ \langle \ln(1 - \lambda_j) \rangle &= \Psi(d_j^*) - \Psi(c_j^* + d_j^*) \\ \langle \ln(\psi_j) \rangle &= \frac{a_j^*}{b_j^*} \\ \langle \lambda_j \rangle &= \frac{c_j}{c_j + d_j} \end{aligned} \quad (\text{A9})$$

where Ψ is Digamma function.

(3) Optimal solution to $Q(\vec{\alpha})$.

$$Q(\vec{\alpha}) = \prod_{j=1}^M \prod_{d=1}^D \mathcal{G}(\alpha_{jd} | u_{jd}^*, v_{jd}^*) \quad (\text{A10})$$

where

$$\begin{aligned} u_{jd}^* &= u_{jd} + \sum_{i=1}^N \langle z_{ij} \rangle \\ v_{jd}^* &= v_{jd} - \sum_{i=1}^N [S_{jd} + \ln(y_{id}) - \langle \beta_{jd} \rangle y_{id}] \langle z_{ij} \rangle \\ S_{jd} &= 1 + \ln(\alpha_{jd}^*) - \frac{1}{\alpha_{jd}^*} - \Psi(\alpha_{jd}^*) + \langle \ln(\beta_{jd}) \rangle \end{aligned} \quad (\text{A11})$$

From the previous equations, we obtain the following expectations:

$$\begin{aligned} \langle \alpha_{jd} \rangle &= \frac{u_{jd}^*}{v_{jd}^*} \\ \langle \ln(\alpha_{jd}) \rangle &= \Psi(u_{jd}^*) - \ln(v_{jd}^*) \end{aligned} \quad (\text{A12})$$

(4) Optimal solution to $Q(\vec{\beta})$.

$$Q(\vec{\beta}) = \prod_{j=1}^M \prod_{d=1}^D \mathcal{G}(\beta_{jd} | s_{jd}^*, t_{jd}^*) \quad (\text{A13})$$

where

$$\begin{aligned}s_{jd}^* &= s_{jd} + \langle \alpha_{jd} \rangle \sum_{i=1}^N \langle z_{ij} \rangle \\ t_{jd}^* &= t_{jd} + \langle \alpha_{jd} \rangle \sum_{i=1}^N \langle z_{ij} \rangle y_{id}\end{aligned}\quad (\text{A14})$$

From the previous equations, we obtain the following expectations:

$$\begin{aligned}\langle \beta_{jd} \rangle &= \frac{s_{jd}^*}{t_{jd}^*} \\ \langle \ln(\beta_{jd}) \rangle &= \Psi(s_{jd}^*) - \ln(t_{jd}^*)\end{aligned}\quad (\text{A15})$$

References

1. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [\[CrossRef\]](#)
2. Lai, Y.; Ping, Y.; He, W.; Wang, B.; Wang, J.; Zhang, X. Variational Bayesian inference for finite inverted Dirichlet mixture model and its application to object detection. *Chin. J. Electron.* **2018**, *27*, 603–610. [\[CrossRef\]](#)
3. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
4. Andrews, J.L.; McNicholas, P.D.; Subedi, S. Model-based classification via mixtures of multivariate t-distributions. *Comput. Stat. Data Anal.* **2011**, *55*, 520–529. [\[CrossRef\]](#)
5. Bouguila, N.; Almakadmeh, K.; Boutemedjet, S. A finite mixture model for simultaneous high-dimensional clustering, localized feature selection and outlier rejection. *Expert Syst. Appl.* **2012**, *39*, 6641–6656. [\[CrossRef\]](#)
6. Elguebaly, T.; Bouguila, N. Background subtraction using finite mixtures of asymmetric Gaussian distributions and shadow detection. *Mach. Vis. Appl.* **2014**, *25*, 1145–1162. [\[CrossRef\]](#)
7. Elguebaly, T.; Bouguila, N. Bayesian Learning of Generalized Gaussian Mixture Models on Biomedical Images. In *Artificial Neural Networks in Pattern Recognition, Proceedings of the 4th IAPR TC3 Workshop, ANNPR 2010, Cairo, Egypt, 11–13 April 2010*; Schwenker, F., Gayar, N.E., Eds.; Springer: Berlin, Germany, 2010; Volume 5998, pp. 207–218. [\[CrossRef\]](#)
8. Lai, Y.; Cao, H.; Luo, L.; Zhang, Y.; Bi, F.; Gui, X.; Ping, Y. Extended variational inference for gamma mixture model in positive vectors modeling. *Neurocomputing* **2021**, *432*, 145–158. [\[CrossRef\]](#)
9. Li, H.; Krylov, V.A.; Fan, P.; Zerubia, J.; Emery, W.J. Unsupervised Learning of Generalized Gamma Mixture Model With Application in Statistical Modeling of High-Resolution SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2153–2170. [\[CrossRef\]](#)
10. Ziou, D.; Bouguila, N. Unsupervised Learning of a Finite Gamma Mixture Using MML: Application to SAR Image Analysis. In *Proceedings of the 17th International Conference on Pattern Recognition, (ICPR 2004), Cambridge, UK, 23–26 August 2004*; pp. 68–71. [\[CrossRef\]](#)
11. Al-Osaimi, F.R.; Bouguila, N. A Finite Gamma Mixture Model-Based Discriminative Learning Frameworks. In *Proceedings of the 14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, 9–11 December 2015*; Li, T., Kurgan, L.A., Palade, V., Goebel, R., Holzinger, A., Verspoor, K., Wani, M.A., Eds.; IEEE: New York, NY, USA, 2015; pp. 819–824. [\[CrossRef\]](#)
12. Beckmann, C.; Woolrich, M.; Smith, S. Gaussian/Gamma mixture modelling of ICA/GLM spatial maps, In *Proceedings of the 9th International Conference on Functional Mapping of the Human Brain*. New York, NY, USA, 19–22 June 2003.
13. Alharithi, F.S.; Almulihi, A.H.; Bourouis, S.; Alroobaea, R.; Bouguila, N. Discriminative Learning Approach Based on Flexible Mixture Model for Medical Data Categorization and Recognition. *Sensors* **2021**, *21*, 2450. [\[CrossRef\]](#)
14. Bourouis, S.; Channoufi, I.; Alroobaea, R.; Rubaiee, S.; Andejany, M.; Bouguila, N. Color object segmentation and tracking using flexible statistical model and level-set. *Multim. Tools Appl.* **2021**, *80*, 5809–5831. [\[CrossRef\]](#)
15. Fan, W.; Bouguila, N.; Bourouis, S.; Laalaoui, Y. Entropy-based variational Bayes learning framework for data clustering. *IET Image Process.* **2018**, *12*, 1762–1772. [\[CrossRef\]](#)
16. Najar, F.; Bourouis, S.; Zaguia, A.; Bouguila, N.; Belghith, S. Unsupervised Human Action Categorization Using a Riemannian Averaged Fixed-Point Learning of Multivariate GGMM. In *Proceedings of the Image Analysis and Recognition—15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, 27–29 June 2018*; pp. 408–415.
17. Ferguson, T.S. Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*; Academic Press: New York, NY, USA, 1983; pp. 287–302.
18. Rasmussen, C.E. A Practical Monte Carlo Implementation of Bayesian Learning. In *Proceedings of the Advances in Neural Information Processing Systems 8, NIPS, Denver, CO, USA, 27–30 November 1995*; Touretzky, D.S., Mozer, M., Hasselmo, M.E., Eds.; MIT Press: Cambridge, MA, USA, 1995; pp. 598–604.

19. Bourouis, S.; Alroobaea, R.; Rubaiee, S.; Andejany, M.; Almansour, F.M.; Bouguila, N. Markov Chain Monte Carlo-Based Bayesian Inference for Learning Finite and Infinite Inverted Beta-Liouville Mixture Models. *IEEE Access* **2021**, *9*, 71170–71183. [\[CrossRef\]](#)
20. Bouguila, N.; Elguebaly, T. A fully Bayesian model based on reversible jump MCMC and finite Beta mixtures for clustering. *Expert Syst. Appl.* **2012**, *39*, 5946–5959. [\[CrossRef\]](#)
21. Jordan, M.I.; Ghahramani, Z.; Jaakkola, T.S.; Saul, L.K. An Introduction to Variational Methods for Graphical Models. *Mach. Learn.* **1999**, *37*, 183–233. [\[CrossRef\]](#)
22. Fan, W.; Bouguila, N. Online Learning of a Dirichlet Process Mixture of Beta-Liouville Distributions Via Variational Inference. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1850–1862. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Elguebaly, T.; Bouguila, N. A Bayesian approach for SAR images segmentation and changes detection. In Proceedings of the 2010 25th Biennial Symposium on Communications, Kingston, ON, Canada, 12–14 May 2010; pp. 24–27. [\[CrossRef\]](#)
24. Zhao, J.; Temimi, M.; Ghedira, H.; Hu, C. Exploring the potential of optical remote sensing for oil spill detection in shallow coastal waters—a case study in the Arabian Gulf. *Opt. Express* **2014**, *22*, 13755–13772. [\[CrossRef\]](#)
25. Singha, S.; Bellerby, T.J.; Trieschmann, O. Detection and classification of oil spill and look-alike spots from SAR imagery using an Artificial Neural Network. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2012, Munich, Germany, 22–27 July 2012; pp. 5630–5633.
26. Brekke, C.; Solberg, A.H. Oil spill detection by satellite remote sensing. *Remote Sens. Environ.* **2005**, *95*, 1–13. [\[CrossRef\]](#)
27. Salberg, A.; Larsen, S.O. Classification of Ocean Surface Slicks in Simulated Hybrid-Polarimetric SAR Data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7062–7073. [\[CrossRef\]](#)
28. Alpers, W.; Holt, B.; Zeng, K. Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sens. Environ.* **2017**, *201*, 133–147. [\[CrossRef\]](#)
29. Solberg, A.H.S.; Storvik, G.; Solberg, R.; Volden, E. Automatic detection of oil spills in ERS SAR images. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1916–1924. [\[CrossRef\]](#)
30. Skrunes, S.; Brekke, C.; Eltoft, T. Characterization of Marine Surface Slicks by Radarsat-2 Multipolarization Features. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5302–5319. [\[CrossRef\]](#)
31. Fingas, M.; Brown, C.E. A Review of Oil Spill Remote Sensing. *Sensors* **2018**, *18*, 91. [\[CrossRef\]](#)
32. Fiscella, B.; Giancaspro, A.; Nirchio, F.; Pavese, P.; Trivero, P. Oil spill detection using marine SAR images. *Int. J. Remote Sens.* **2000**, *21*, 3561–3566. [\[CrossRef\]](#)
33. Gambardella, A.; Giacinto, G.; Migliaccio, M.; Montali, A. One-class classification for oil spill detection. *Pattern Anal. Appl.* **2010**, *13*, 349–366. [\[CrossRef\]](#)
34. Topouzelis, K.; Karathanassi, V.; Pavlakis, P.; Rokos, D. Detection and discrimination between oil spills and look-alike phenomena through neural networks. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 264–270. [\[CrossRef\]](#)
35. Karantzas, K.; Argialas, D. Automatic detection and tracking of oil spills in SAR imagery with level set segmentation. *Int. J. Remote Sens.* **2008**, *29*, 6281–6296. [\[CrossRef\]](#)
36. Chang, L.; Tang, Z.S.; Chang, S.H.; Chang, Y. A region-based GLRT detection of oil spills in SAR images. *Pattern Recognit. Lett.* **2008**, *29*, 1915–1923. [\[CrossRef\]](#)
37. Solberg, A.H.S.; Brekke, C.; Husoy, P.O. Oil Spill Detection in Radarsat and Envisat SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 746–755. [\[CrossRef\]](#)
38. Keramitsoglou, I.; Cartalis, C.; Kiranoudis, C.T. Automatic identification of oil spills on satellite images. *Environ. Model. Softw.* **2006**, *21*, 640–652. [\[CrossRef\]](#)
39. Cantorna, D.; Dafonte, C.; Iglesias, A.; Varela, B.A. Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms. *Appl. Soft Comput.* **2019**, *84*, 105716. [\[CrossRef\]](#)
40. Krestenitis, M.; Orfanidis, G.; Ioannidis, K.; Avgerinakis, K.; Vrochidis, S.; Kompatsiaris, I. Oil Spill Identification from Satellite Images Using Deep Neural Networks. *Remote Sens.* **2019**, *11*, 1762. [\[CrossRef\]](#)
41. Orfanidis, G.; Ioannidis, K.; Avgerinakis, K.; Vrochidis, S.; Kompatsiaris, I. A Deep Neural Network for Oil Spill Semantic Segmentation in Sar Images. In Proceedings of the 2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, 7–10 October 2018; pp. 3773–3777.
42. Song, D.; Ding, Y.; Li, X.; Zhang, B.; Xu, M. Ocean Oil Spill Classification with RADARSAT-2 SAR Based on an Optimized Wavelet Neural Network. *Remote Sens.* **2017**, *9*, 799. [\[CrossRef\]](#)
43. Li, J.; Du, Q.; Li, Y. An efficient radial basis function neural network for hyperspectral remote sensing image classification. *Soft Comput.* **2016**, *20*, 4753–4759. [\[CrossRef\]](#)
44. Shaban, M.; Salim, R.; Abu Khalifeh, H.; Khelifi, A.; Shalaby, A.; El-Mashad, S.; Mahmoud, A.; Ghazal, M.; El-Baz, A. A Deep-Learning Framework for the Detection of Oil Spills from SAR Data. *Sensors* **2021**, *21*, 2351. [\[CrossRef\]](#)
45. Zeng, K.; Wang, Y. A Deep Convolutional Neural Network for Oil Spill Detection from Spaceborne SAR Images. *Remote Sens.* **2020**, *12*, 1015. [\[CrossRef\]](#)
46. Topouzelis, K.N. Oil Spill Detection by SAR Images: Dark Formation Detection, Feature Extraction and Classification Algorithms. *Sensors* **2008**, *8*, 6642–6659. [\[CrossRef\]](#)
47. Teh, Y.W. Dirichlet Process. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 280–287.
48. Sethuraman, J. A constructive definition of Dirichlet priors. *Stat. Sin.* **1994**, *4*, 639–650.

49. Blei, D.M.; Jordan, M.I. Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **2006**, *1*, 121–143. [[CrossRef](#)]
50. Ishwaran, H.; James, L.F. Gibbs sampling methods for stick-breaking priors. *J. Am. Stat. Assoc.* **2001**, *96*, 161–173. [[CrossRef](#)]
51. Oppor, M.; Saad, D. *Advanced Mean Field Methods: Theory and Practice*; MIT Press: Cambridge, MA, USA, 2001.
52. Blei, D.M.; Jordan, M.I. Variational methods for the Dirichlet process. In *Machine Learning, Proceedings of the Twenty-First International Conference (ICML 2004), Banff, AL, Canada, 4–8 July 2004*; Brodley, C.E., Ed.; ACM International Conference Proceeding Series; ACM: New York, NY, USA, 2004; Volume 69.
53. Sato, M. Online Model Selection Based on the Variational Bayes. *Neural Comput.* **2001**, *13*, 1649–1681. [[CrossRef](#)]
54. Fan, W.; Bouguila, N. Online variational learning of generalized Dirichlet mixture models with feature selection. *Neurocomputing* **2014**, *126*, 166–179. [[CrossRef](#)]
55. Hoffman, M.D.; Blei, D.M.; Bach, F.R. Online Learning for Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: New York, NY, USA, 2010; pp. 856–864.
56. Manouchehri, N.; Nguyen, H.; Koochemeshkian, P.; Bouguila, N.; Fan, W. Online Variational Learning of Dirichlet Process Mixtures of Scaled Dirichlet Distributions. *Inf. Syst. Front.* **2020**, *22*, 1085–1093. [[CrossRef](#)]
57. Konik, M.; Bradtke, K. Object-oriented approach to oil spill detection using ENVISAT ASAR images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *118*, 37–52. [[CrossRef](#)]
58. Topouzelis, K.; Psyllos, A. Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *68*, 135–143. [[CrossRef](#)]
59. Wang, C.; Mouche, A.; Tandeo, P.; Stopa, J.E.; Longépé, N.; Erhard, G.; Foster, R.C.; Vandemark, D.; Chapron, B. A labelled ocean SAR imagery dataset of ten geophysical phenomena from Sentinel-1 wave mode. *Geosci. Data J.* **2019**, *6*, 105–115. [[CrossRef](#)]
60. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016*; IEEE Computer Society: Washington, DC, USA, 2016; pp. 770–778.
61. Topouzelis, K.; Stathakis, D.; Karathanassi, V. Investigation of genetic algorithms contribution to feature selection for oil spill detection. *Int. J. Remote Sens.* **2009**, *30*, 611–625. [[CrossRef](#)]
62. Ferraro, G.; Pavlakakis, P.; Tarchi, D.; Sieber, A.; Ferraro, G.; Vincent, G. *On the Monitoring of Illicit Discharges—A Reconnaissance Study in the Mediterranean Sea*; EUR 19906 EN; European Commission: Brussels, Belgium, 2001.
63. Chatziantoniou, A.; Karagaitanakis, A.; Bakopoulos, V.; Papandroulakis, N.; Topouzelis, K. Detection of Biogenic Oil Films near Aquaculture Sites Using Sentinel-1 and Sentinel-2 Satellite Images. *Remote Sens.* **2021**, *13*, 1737. [[CrossRef](#)]
64. Chatziantoniou, A.; Bakopoulos, V.; Papandroulakis, N.; Topouzelis, K. Detection of biogenic oil film near aquaculture sites seen by Sentinel-2 multispectral images. In *Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions 2020*; International Society for Optics and Photonics: San Diego, CA, USA, 2020; p. 4. [[CrossRef](#)]