GUIDED MCMC FOR SPARSE BAYESIAN MODELS TO DETECT RARE EVENTS IN IMAGES SANS LABELED DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Detection of rare events in images is a challenging task because of two main problems, the first problem is the lack of labeled data for rare category class and the second problem is a highly imbalanced data problem. Training models in this scenario becomes hard. Unsupervised methods do not apply as we need to detect rare events automatically. Rule-based methods seem to be the only viable solution, but it is tedious to come up with a set of rules covering all corner cases. Even the recently popular zero-shot learning techniques required to be pre-trained on auxiliary datasets. In the given scenario, we propose an approach to provide little guidance from experts as an input into a hierarchical Bayesian model. The guidance influences the Markov chain Monte Carlo (MCMC) based inference technique of the model. After the steady-state is obtained for the underlying Markov chain, it is possible to compute the posterior probability of the presence of the rare event in a given image. The proposed method neither needs any labeled data nor required pre-training, unlike zero-shot learning. The proposed technique has been observed to outperform the state-of-the-art unsupervised image classification techniques.

1 INTRODUCTION

The significance of rare events is that they occur rarely but their occurrence can have a high impact. Heavy rainfall to solar bursts (White, 2007) there are plenty of such phenomena in real life which belong to the category of rare events. In this paper, we focus on rare events present in a collection of images.

Detecting rare events in images is hard. Rare events detection falls into the classification problem where convolutional neural networks (CNNs) are the de-facto standard. However, CNNs do need some amount significant amount of examples to train the parameters. However, the presence of rare events is not only insignificant but also relevant to specific problems. For example, pedestrians crossing a road without signals. In the problem of rare event detection, the focus is on detecting the rare events instead of overall accuracy. Supervised techniques do not naturally fit in this case. Unsupervised methods are not applicable as in the end we want rare events to be labeled automatically. One can resort to rule-based techniques, but as always it is hard to design an exhaustive set of rules especially when rare events are critical to detect.

In this paper, we propose a novel mechanism Guided MCMC for Bayesian Models (**GMBM**) to detect rare events without using any labeled data or pre-training using auxiliary datatsets. Our method is more based on the intuition of rule-based systems and extending them through machine learning mechanism. As in rule-based systems, we collect some guidance from human experts who are aware of the dataset and the problem. Then we feed this guidance into the inference process of a hierarchical Bayesian model as in machine learning to compute the posterior of the presence of rare events in a given image.

Hierarchical Bayesian models are well-known techniques to build distributions suitable to model some type of observations. Latent Dirichlet allocation (LDA) (Blei et al., 2003) is one such example to model bag-of-words of documents. Then we learn the distribution through estimating the parameters of the hierarchical Bayesian model. In general, hierarchical Bayesian models involve latent random variables and often inference becomes intractable. Markov chain Monte Carlo

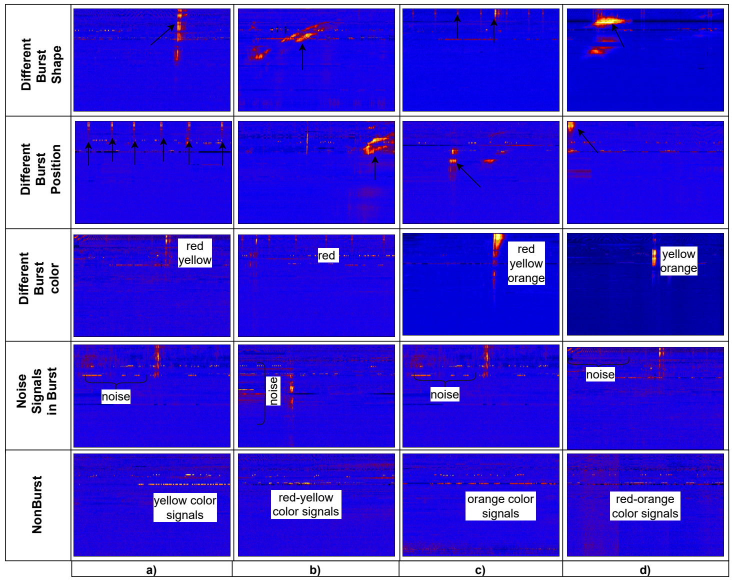

Figure 1: **Illustration of rare event in images using solar burst and associated challenges to detect them.** In the first row, we can see that there are different shapes of bursts. In the second row, we can see that positions of bursts are not fixed. In the third row, we can see that even color patches of bursts are not the same. In the fourth row, we can observe that noise signals are also present in burst images themselves. In the last row we can see Non-Burst signals but have similarity with bursts images.

(MCMC) (Geyer, 1992) is a well-established technique to infer and learn such models. Due to the non-convexity of the underlying optimization problem, the Markov chain behind the MCMC technique can converge to some local optima. Once we achieve a steady state of the underlying Markov chain we can compute the posterior of the latent random variables. Depending on the design of the hierarchical Bayesian models, we can infer some decisions through the posterior of such latent variables.

We design a hierarchical Bayesian model, and then use MCMC to infer whether a rare event is present or not. However, here comes our novelty that, instead of running vanilla MCMC, we guide the Markov chain with appropriate input from human experts so that we can infer the presence of rare events without using any label. Secondly, we use sparsity along with mixture of distributions in such a way that gives importance to rare events.

The Contributions of our work are as follows:-

- We have proposed a novel mechanism based on MCMC for hierarchical Bayesian models to detect rare events in images without using any labeled data. The mechanism accepts input from humans to guide MCMC to decide the label.
- The hierarchical Bayesian model is designed with sparsity and mixture of distributions in such a way that it gives importance to rare events much in the spirit of importance sampling.
- We have demonstrated the efficacy of our model in detecting rare events in images outperforming state-of-the-art unsupervised image classification techniques.

2 Problem Definition: Detecting Rare Events in Images sans Labels

Let $\mathcal{D} = \{x_i\}_{i=1}^{n}$ denotes a collection n images. Each image x_i consists of m patches $\{x_{ij}\}_{j=1}^{m}$, where a patch is a collection of spatially adjacent pixels. The task is to identify if a rare event is present in an image, i.e. set $y_i = 1$ if x_i contains a rare event, otherwise $y_i = 0$. To be precise, we denote an event $\mathcal{E}$ by its visual characteristics $\{\mathcal{E}_a\}$. We say an event $\mathcal{E} \in x_i$ for some i, if $\{\mathcal{E}_a\}$

is present in the patches of x_i. We further designate an event $\mathcal{E}$ to be *rare* if $P(\mathcal{E} \in x_i) < \epsilon < 1$, where ϵ is a small quantity and generally prescribed the domain expert.

Let $\mathcal{R}$ denotes the set of rare events for a given application and dataset $\mathcal{D}$. The set $\mathcal{R}$ is generally defined by the domain experts along with the visual characteristics. Then the task is to set $y_i = 1$ if $\mathcal{E} \in x_i$ such that $\mathcal{E} \in \mathcal{R}$. In other words $y_i = 1$ if $P(\mathcal{E}|x_i) > \delta$, where $\delta > 0$ is the threshold. We can form the task as set $y_i = 1, \text{ if } \quad P(\mathcal{E}|\{x_{ij}\}_{j=1}^{m}) > \delta$. Let $\mathcal{V}$ denotes the vocabulary or the set of visual words representing the visual characteristics in such a way that each $x_{ij} \in \mathcal{V}$ and $\mathcal{E}_a \in \mathcal{V}$. Then we can further refine the objective is to compute $P(\{\mathcal{E}_a\}|\{x_{ij}\}_{j=1}^{m})$.

2.1 Example: Solar burst detection

This work is largely motivated by the problem of detecting solar bursts. In the solar radio observations, solar bursts are considered to be as rare events. Solar radio bursts (Monstein, 2011) are classified largely based on how they appear in dynamic spectrum observations from radio spectrographs. In Aug 2021, the number of bursts observed was 149 out of total reported observations 678 taken from 55 active stations[1]. The shapes of solar bursts can be seen as big red spikes or harmonic shapes and typically their sizes are bigger than noise signals which can be seen in Figure 1.

2.2 Challenges for detecting rare events in images

It is hard to apply supervised techniques in rare event detection are: (i) lack of labeled data for training, (ii) highly imbalanced data. One can not use unsupervised methods because the task is more of automatic classification and in turn detection rather than pattern analysis.

Observing Figure 1, it may appear that the pattern of solar bursts can be captured through rules. The same thing might be done for other rare events in images. However, the challenge in such methods is always to design rules carefully exhaustively covering corner cases. This is extremely hard and often the sole reason of opting the data-driven methods.

In order to explore this aspect more closely, we have developed our own rule-based heuristic method to detect solar bursts. The algorithm is presented in Algorithm 2 in the appendix. The basic idea of the method is to use contour and color to detect bursts. The domain experts can prescribe the color and shape to some extent and depending on that we can define a rule, such as, if some area is of red color and greater than a certain threshold but also not too big, then we can label that as burst.

Although it is very easy for humans to use such rules as the slight modifications in rules on a case-by-case basis can be very easily adapted by humans due to their domain knowledge. However, it is infeasible to capture even slight variations in color and contour for a rule. The second challenge is that sometimes noise signals are also a similar size to solar Bursts due to which it will end up predicting the image as a Burst image but in reality, it's a Non-Burst image. Some more challenges for solar burst detection can be seen in Figure 1.

3 Related Works

Hamaguchi et al. (2019) have proposed a novel method for rare event detection from an image pair with the class imbalanced dataset. Their idea is that to learn disentangled representations from only low-cost negative samples. This paper is the most recent work in the area of rare events detection in images. But in our case, we are not using any training data even if data belongs to negative samples. Recently for solar radio Burst classification Chen et al. (2017) tried to use the CNN model for performing supervised Burst image classification.

In the area of unsupervised image classification, recently Van Gansbeke et al. (2020) designed SCAN model which have shown astonishing performance in unsupervised image classification. Park et al. (2021) studied to build an innovative RUC model which can be used as an add-on-module on top of models like SCAN to improve classification accuracy. Dang et al. (2021) proposed a method called Nearest Neighbor Matching (NNM) to match samples with their nearest neighbors

[1] http://www.e-callisto.org/Data/data.html

from both local and global levels and has shown that superior unsupervised classification performance against state-of-art methods.

Hingmire & Chakraborti (2014) propose a weakly supervised algorithm in which supervision comes in the form of labeling of Latent Dirichlet Allocation (LDA) topics. They have used this approach to perform text classification. Xian et al. (2018) gave a detailed description about Zero-shot learning where its main aim is to recognize objects whose instances may not have been seen during training. Although this work will not be strongly applicable to our case as we are not using any training data. Zero-shot learning is also closely related to transfer learning. Similarly, Wang et al. (2020) write about a Few-shot learning survey paper where it can learn new tasks containing only a few labels with supervised information by incorporating prior knowledge but the core issue of Few-shot learning is the unreliable empirical risk minimizer that makes Few-shot learning hard to learn.

4 Proposed Method: Guided MCMC for Sparse Bayesian Models

First, we describe the basic setup how we can leverage hierarchical Bayesian models and MCMC for the task of classification. Then we will develop on top of the basic setup by introducing two key modifications: (i) sparsity and focus to give importance to rare events, and (ii) guidance to MCMC to connect posterior of a latent variable to a label in case we do not have labels. We will refer to the proposed method as GMBM.

4.1 MCMC for Bayesian models for classification

A wide range of hierarchical Bayesian models can be represented as follows:

$$\prod_{i=1}^{N} P(\theta_i|\alpha) \ \Big(\prod_{j=1}^{M} P(x_{ij}|z_{ij},\beta)P(z_{ij}|\theta_i)\Big), \tag{1}$$

where $\{z_{ij}\}, \{\theta_i\}$ are the latent random variables in two different levels of the hierarchy. $\{z_{ij}\}$ are the local variables corresponding each observed unit in $\{x_{ij}\}$, whereas θ_i is common across x_i. α, β are the parameters of the model which we want to learn. We can represent latent Dirichlet allocation (LDA) and several topic models in this format. For our case $\{x_{ij}\}$ are image patches that come from the global vocabulary $\mathcal{V}$.

Given this basic setup, we set θ_i to be a K dimensional vector such that $\sum_{k=1}^{K}\theta_{ik} = 1, \theta_{ik} \geq 0, \ \forall k$. We can model $P(\theta_i|\alpha)$ as $Dirichlet(\alpha 1_K)$. In our case, K is the number of events where each event belongs to a particular class and the label of image x_i is y_i. As we are more concerned about the presence of rare event, so following the setup in Sec. 2, we denote $y_i = 1$ if $\hat{\theta}_{ir} > \delta$, where $\hat{\theta}_i$ is the posterior of θ_i given image x_i, r is the index of the rare event and $\delta > 0$ is the threshold. Then the task is to compute the posterior probability $P(\theta_i|\{x_{ij}\}_{j=1}^{m})$ by applying Bayes theorem.

Specification of other variables can be dependent on the datasets and applications and do not interfere with the classification setup as posed here. Due to the nature of the hierarchy in Eq 1, computing $P(\theta_i|x_i)$ becomes hard and we generally resort to MCMC or Gibbs sampling. In the case of conjugacy, we can even collapse some of the variables leading to faster convergence (Porteous et al., 2008).

4.2 Sparsity and Importance for Rare Events

Note that in our case there is a chance that some rare event $\mathcal{E}$ is present in the image patches $\{x_{ij}\}$. The event $\mathcal{E}$ is manifested in the image through visual words $\{\mathcal{E}_a\}$ among visual words in the patches $\{x_{ij}\}$ such that $P(\mathcal{E} \in \{x_{ij}\}) < \epsilon$. One key challenge is that ϵ can be very small. Hence unless we put specific care it is difficult to catch rare events through statistical models.

To be more precise $P(\hat{\theta}_{ir} > \delta)$ can be arbitrarily small and may be difficult to distinguish with noise and false positives if we keep δ too small even if $\mathcal{E} \in x_i$. This is due to the fact that θ_i captures the proportion of all events in image x_i, and along with rare events there are other normal

Algorithm 1 Generative process of the proposed Bayesian model

for visual-words v=1,2,...,V **do**:
  sample visual-word selection probability $\psi_v \sim Beta(1, \varrho)$
end for
for events k=1,2,...,K **do**:
  for visual-words 1=1,2,...,V **do**:
    sample visual-word selection probability $\phi_{kv} \sim Bernoulli(\psi_v)$
  end for
  draw event $\beta_k \sim Dirichlet(\kappa 1_V.\phi_k)$
  event selection probability $\pi_k \sim Beta(1, \varrho)$
end for
for images i = 1 to N **do**:
  sample distribution over events proportions $\eta_i \sim Dirichlet(\zeta 1_H)$
  for events proportions $h = 1, ..., H$ **do**:
    for events k=1,2,...,K **do**:
      sample $\xi_{ihk} \sim Bernoulli(\pi_k)$
    end for
    sample $\theta_{ih} \sim Dirichlet(\alpha 1_K.\xi_{ih})$
  end for
  for patches $j = 1, ..., M$ **do**:
    sample event proportion $g_{ij} \sim mult(\eta_i)$
    sample event $z_{ij} \sim mult(\theta_{i,g_{ij}})$
    sample patch $x_{ij} \sim mult(\beta_{z_{ij}})$
  end for
end for

events in x_i. Normal events being prominently present in the collection as well as in the image, $\theta_{ir} = 1 - \sum_{k=1,\neq r}^{K} \theta_{ik}$ can be insignificant.

In order to mitigate this issue our proposal is to give *importance* to rare events. Our proposal is in the spirit of *importance sampling*. We allow the model to have an events proportion vector that puts zero weight on most of the events and non-zero to a few including the rare event mainly. Now this proportion vector being a probability mass function, non-zero entries get boosted up. Using the technique of sparsity we put zero weights to many common events. In this way, we provide *importance* to the rare events. Then similar to importance sampling we down-scale the proposal distribution. This we achieve by putting a distribution over the proportion vectors.

4.2.1 Additional proposal distribution for rare events

Although a proportion vector is sufficient to model all the events in an image, we propose to use additional proportion vectors such as instead of θ_i for x_i, now we have $\{\theta_{ih}\}_{h=1}^{H}$. We want one of the proportion vectors to put focus on the rare event. So we induce sparsity in the proportion vectors. If $\theta_{ih} \sim Dirichlet(\alpha 1_K)$, then we sample binary random variables $\xi_{ihk} \sim Bernoulli(\pi_k)$, and enforce sparsity in the proportion vectors as follows:

$$\sum_{k=1}^{K} \theta_{ihk}\, \xi_{ihk} = 1, \forall i, h; \quad \theta_{ihk}\, \xi_{ihk} \geq 0, \;\; \xi_{ihk} \in \{0, 1\}, \;\; \forall k, \tag{2}$$

where $a.b$ denotes element wise product of two vectors of the same length. The above construction can be represented as $\theta_{ih} \sim Dirichlet(\alpha 1_K.\xi_{ih})$. Similarly we induce sparsity in the event vectors that is proportion over visual words for an event:

$$\sum_{v=1}^{V} \beta_{kv}\, \phi_{kv} = 1, \forall k; \quad \beta_{kv}\, \phi_{kv} \geq 0, \;\; \phi_{kv} \in \{0, 1\}, \;\; \forall k, \tag{3}$$

or alternatively $\beta_k \sim Dirichlet(\kappa 1_V.\phi_k)$. Thus not all events put weight on all visual words, and effectively events are more likely to be represented by distinguishing visual words.

4.3 Proposed hierarchical Bayesian model

Depending on the above constructions, we define the hierarchical Bayesian model as below:

$$\prod_{i=1}^{N}\left(P(\eta_i|\zeta)\prod_{h=1}^{H}P(\theta_{ih}|\alpha)\right)\left(\prod_{j=1}^{M}P(x_{ij}|z_{ij},\beta)P(z_{ij}|\{\theta_{ih}\},\{\xi_{ih}\},g_{ij})P(g_{ij}|\eta_i)\right), \tag{4}$$

where z_{ij} denotes the index of event to draw the patch x_{ij}, g_{ij} is the index of the proportion vector over events $\{\theta_{ih}\}$. Thus, g_{ij} acts like a selection random variable and can be sampled from a discrete distribution with parameter η_i. The rest of the model is quite similar to the topic models. The complete generative process is given in Algorithm 1. The model has some resemblance with the subtle topic models (STM) Das et al. (2013), where they have also used multiple proportions over topics in a document to discover subtly manifested software concerns.

4.3.1 Posterior Inference with Sparsity and Importance

We use the collapsed Gibbs sampling approach (Porteous et al., 2008) to sample the latent variables using the posterior conditional distribution. We will use the notation for counts as follows. i is the image index, k is the event index and j is the image patch position index. n represents the counting variable and indices are put in the subscript, where "." represents marginalization. ϕ_{kv} denotes a binary value of visual word v index by event k. Inference of ξ and g are required to assign events. The notation table can be found in the appendix A.1. The detailed sampling of ξ and g can be found in the appendix A.2. Mapping between visual words and patches is mentioned in the appendix section A.4. The conditional probability of event assignment of patch j at event k in image i can be expressed as:

$$\begin{aligned} p(z_{ij}=k|x,z_{-ij}) &\propto p(x_{ij}|z_{ij}=k,z_{-ij},\phi,\kappa)p(z_{ij}=k|z_{i,-j},g_i,\xi_i,\alpha) \\ &= \underbrace{\frac{\phi_{kj}\kappa + n^{-ij}_{...kj}}{\sum_v \phi_{kv}\kappa + n^{-ij}_{...k}}}_{\beta}\ \underbrace{\frac{\xi_{ig_{ij}k}\alpha + n^{-ij}_{i..g_{ij}k}}{\sum_k \xi_{ig_{ij}k}\alpha + n^{-ij}_{i..g_{ij}}}}_{\theta} \end{aligned} \tag{5}$$

where β is a event-visual word distribution of size $K \times V$ and θ is an image-event distribution of size $N \times H \times K$.

4.4 Guided MCMC for Rare Event Detection

Recall that, our objective is to infer $\hat{\theta}_i = P(\theta_i|\{x_{ij}\})$. Here utilizing the conjugacy between Dirichlet and multinomial, we can marginalize out $\{\theta_i\}$ and other continuous random variables and build the Markov chain only the space of discrete random variables such as $\{z_{ij}\}$. This scheme is called collapsed sampling and in general lead to faster convergence. Then after the convergence is achieved, we can retrieve back $\{\theta_i\}$ and other continuous random variables again utilizing the conjugacy property between Dirichlet and multinomial, that is if $\theta_i \sim Dirichlet(\alpha 1_K)$, then

$$\hat{\theta}_i \sim Dirichlet(\alpha + c_{i1}, \alpha + c_{i2}, \ldots, \alpha + c_{iK}), \ \ c_{il} = \sum_{j=1}^{m} I[z_{ij} = l], \tag{6}$$

where $I[]$ is the indicator function. Note that, as $P(z_{ij} = l) = \theta_{il}$, the index in the proportion vector is meaningful. That is, if r is the index of the rare event $\mathcal{E}$, then $P(z_{ij} = r) = \theta_{ir}$. Further notice that, $z_{ij} = r$ means that x_{ij} belongs to the rare event. In other words rare event gets manifested in the ith image. Therefore, we can say that $y_i = 1$ if $\hat{\theta}_{ir} > \delta$ if we know that r is the index of the rare event.

Here comes our key contribution, that instead of starting the Markov chain at a random initial point, we start at a point where we set r as the index of the rare event a priori. Due to the structure of the model, the position remains unchanged throughout the MCMC procedure. Hence, after the convergence when we get $\hat{\theta}_{ir}$ we can easily get the value of y_i or label x_i if a rare event is present or not. We call this as guided MCMC and notice that, we do not use any label information throughout the learning process to classify images.

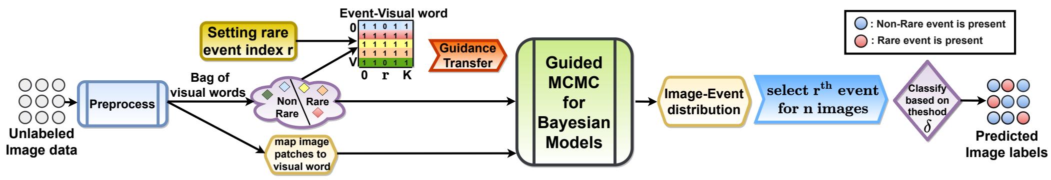

Figure 2: The working pipeline of our proposed GMBM model. The guidance is provided to MCMC for Bayesian Models that give importance to rare events in images while learning the image-event distribution.

We are providing guidance to the sampling process to detect rare events. Recall that, the idea is to fix the index of the rare event as r a priori. Then we start the Markov chain with that information. We do that by carefully initializing the Markov chain instead of random initialization. We will set event r as a rare event index in $\{\phi_k\}_{k=1}^{K}$. It can be expressed as follows:

$$\forall v \in \mathcal{V}, \ \phi_{rv} = \begin{cases} 1, & \text{if } \ v \ \text{ is a rare visual word} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

For non-rare event indexes, we have two ways to fill ϕ_{kv}, first case is either we can put 1 for all entries, or the second case is to put 1 only in the place of non-rare visual words indexes. After setting up the guidance ϕ_{kv} this will remain fixed throughout the finding of posterior inference.

4.5 Decision Process and Analysis

The decision process is described in the Algorithm 3. In order to ensure the method to work, we have to ensure two things: (i) the rare event index as prescribed in the beginning remains the same at the time of making the decision, (ii) we can correctly assign weight for the rare events in one of the proportion vectors.

We need to check $P(\mathcal{E}|x_i)$ and if it is above a threshold δ we claim the presence of rare event in the ith image. Given the hierarchical Bayesian modeling setup in Eq. 4, we use $\hat{\theta}_i$ as the representation for the ith image. We compute $\hat{\theta}_i$ using Eq. 6 and Eq. 5. Now from Eq. 5 note that the index of the event is passed across the iterations through the count variables which are global throughout the lifetime of the sampling process starting from the initialization. Hence, the index of the rare event once set to r will remain at r even after the chain converges or iteration terminates. Given that, we only need to ensure that if a rare event is present then we are able to capture that, and we can compute that as follows:

$$P(\mathcal{E} \in x_i) \approx P(\mathcal{E} \in \{\hat{\theta}_{ih}\}) = \sum_{h=1}^{H} \hat{\theta}_{ihr} \tag{8}$$

Note that $P(z_{ij} = r) = 0$ if $g_{ij} = h$ and $\xi_{ihr} = 0$. Due to the same reason it is possible that $g_{ij} = h$, $\xi_{ihk} = 0$ for all $k \neq r$, then $P(z_{ij} = r) = 1$. Thus even if r denotes the index of rare event it can have high probability. In such cases, other proportion vectors $\{\theta_{il}, \ l \neq h\}$ can model the normal events. We have also experimentally shown that the event-visual words ϕ is helping to find rare event in event-visual words distribution in Table 6, 7, and 8 respectively under the appendix section. The working pipeline diagram of our proposed GMBM model can be seen in Figure 2.

We can analyse the method from a different perspective as follows. Now since we have set r^{th} event as a rare event in ϕ_{kv} then to determine the conditional probability of x_{ij} patch given event r^{th} is as follows:

$$p(x_{ij}|z_{ij} = r, z_{-ij}, \phi, \kappa) = \int d\beta \, p(x_{ij}|z_{ij} = r, \beta) p(\beta|x_{-ij}, z_{-ij}, \phi, \kappa) = \frac{\phi_{rv}\kappa + n_{\ldots rv}^{-ij}}{\sum_u \phi_{ru}\kappa + n_{\ldots r}^{-ij}}.$$

Notice that, if $\phi_{rv} = 0$ in setting the guidance, x_{ij} can not contain visual word v attributing to event r. Whereas, if $\phi_{rv} = 1$ i.e. v is a characteristic visual word for the rare event, then x_{ij} containing the visual word v can be claimed to contain the rare event.

Table 1: Summary of datasets.

Image Datasets	Burst Dataset	Synthetic Burst Datasets(4)	Fruit Datasets(2)	Aeroplane-sky_DS
Classes	Burst(B), Non-Burst(NB)	Burst(B), Non-Burst(NB)	Huckleberry(Hu), Red Apple(Ap), Banana(Ba), Watermelon(Wa)	Aeroplane(A), Sky(S)
Class Counts	{B:50 images, NB:50 images}	DS_1{B:475, NB:525} DS_2{B:259, NB:741} DS_3{B:90, NB:910} DS_4{B:25, NB:975}	DS_1{Hu:490, Ap:492, Ba:490, Wa:475} DS_2{Hu:50, Ap:492, Ba:490, Wa:475}	{A:25,S:475}

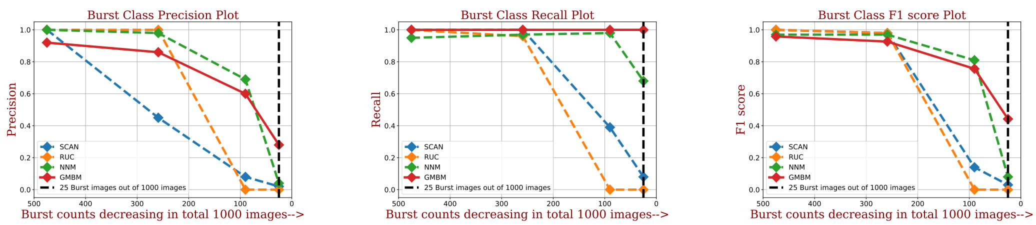

Figure 3: Summary of Precision, Recall, F1 score on four synthetic Burst datasets of Burst class. Here we have shown that how all models are performing when Burst class ratio is reduced from 50% to 2.5% in a total of 1000 images. As the target class becomes very rare, GMBM becomes way better than the other models.

5 EXPERIMENTS

Baselines. We have chosen current state-of-art unsupervised image classification methods which can work without labels. We have used the SCAN model as our first baseline, RUC model as our second baseline, NNM model as our third baseline. The working of SCAN, RUC, and NNM models is already discussed under the related work section 3. We have also used our own heuristic method as a baseline for the original Burst dataset.

Datasets and Experimental Settings. We have used burst dataset[2], four synthetic burst datasets, two fruit datasets using famous fruits360 dataset (Mureșan & Oltean, 2018) and lastly aeroplane-sky dataset made from pascal-voc 2007 dataset (Everingham et al., 2007) and synthetically generated 475 blue sky images. The summary of all datasets can be found in Table 1. The detailed generative process for synthetic burst can be found in Algorithm 5. Dataset descriptions in detail can be found in the appendix section A.5. The detailed experimental setup of our model and baselines are mentioned in the appendix section A.6.

Guidance in the proposed GMBM technique. In the burst dataset, the guidance for Non-Burst event index is given by filling 1 in all indexes and for Burst event index, we have filled 1 only in place of rare visual words indexes i.e red, yellow, and orange color indexes. In the fruit dataset, the guidance is given for each fruit for example in huckleberry event index we have set 1 for blue color index as rare visual word and rest we have filled 0 in all color indexes of huckleberry event index. Similarly for other fruits we did the same procedure. In the aeroplane-sky dataset, the guidance for sky-event is given by filling 1 in non-rare visual word index i.e. sky blue color and for aeroplane event, we have filled 1 in all rare visual words indexes i.e white, black, and grey. More detailed in guidance part can be found in Table 6, 7, and 8 respectively under the appendix section.

5.1 RESULTS

Detecting bursts as rare events. As we have discussed, this work is motivated with the problem of detection of solar bursts which is considered as a hard problem, we compare the proposed technique GMBM with the state of the art methods in that task. Table 2 shows that GMBM is significantly superior.

[2]source: http://rac.ncra.tifr.res.in/da/oc/oecda.html

Table 2: Comparison on detection of Solar bursts as rare events. GMBM performed best among all models. Even our heuristic method based on color contour detection is not far behind.

	Burst class			
Method	**Precision**	**Recall**	**F1 score**	**Accuracy**
Color Contour Detection	0.74	0.88	0.80	79%
SCAN	0.70	0.32	0.44	59%
RUC	0.65	**0.90**	0.76	71%
NNM	0.93	0.54	0.68	75%
GMBM	**0.97**	**0.90**	**0.93**	**94%**

Table 3: Classification performance on fruits dataset when Huckleberry class is made as a rare class. Precision, Recall, and F1 score are mentioned for the Huckleberry class and Overall Accuracy is mentioned for all four classes.

Dataset	Class	Model	Precision	Recall	F1 Score	Accuracy
Fruit_DS_1 (balanced dataset)	**Huckleberry** (490 images)	**SCAN**	**1.00**	**1.00**	**1.00**	74%
		RUC	**1.00**	**1.00**	**1.00**	74%
		NNM	**1.00**	0.99	0.99	**99%**
		GMBM	**1.00**	**1.00**	**1.00**	83%
Fruit_DS_2 (imbalanced dataset)	**Huckleberry** (50 images)	**SCAN**	0.18	0.9	0.29	62%
		RUC	0.00	0.00	0.00	44%
		NNM	0.31	0.78	0.45	**90%**
		GMBM	**1.00**	**1.00**	**1.00**	78%

Table 4: Classification performance on Aeroplane-Sky dataset. Aeroplane is the rare event in this case.

	Aeroplane in sky			Sky			
Model	**Precision**	**Recall**	**F1 Score**	**Precision**	**Recall**	**F1 Score**	**Accuracy**
SCAN	0.14	0.80	0.24	0.99	0.74	0.84	74.0%
RUC	**1.00**	0.84	0.91	0.99	**1.00**	**1.00**	99.2%
NNM	0.40	**1.00**	0.57	**1.00**	0.92	0.96	92.6%
GMBM	**1.00**	0.96	**0.98**	**1.00**	**1.00**	**1.00**	**99.8%**

Varying the Degree of Rarity. We wanted to compare GMBM with the state of the art by varying the percentage of burst images in dataset. For this purpose, we have synthetically (see Algorithm 5) created solar dataset with varied number of bursts. Figure 3 shows that as the degree of rarity increases GMBM outperforms the state of the art methods.

Fruit and Aeroplane-Sky Datasets. We have also experimented on simple datasets like fruits and aeroplane-sky datasets to perform unsupervised image classification. The results are shown in Table 3 and Table 4 respectively. These two experiment shows that our model can be applied to a similar kind of above datasets for the task of unsupervised image classification.

It can be observed that GMBM is particularly better than other models in case of detecting rare events, as in the case of Huckleberry class in Table 3. Aeroplane being the rare event, GMBM gives very high F1 score compared to the baselines.

6 Conclusion

We have proposed an unsupervised method to detect rare events in images by giving guidance to MCMC for Bayesian models. The proposed model GMBM is outperforming the current state-of-art unsupervised image classification in terms of F1 score for a rare class as well as we have also shown that our model can be applied to simple datasets. For future work, we like to explore how to make optimal visual words to make our model more generalize for any image dataset to detect rare events.

REFERENCES

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

Sisi Chen, Long Xu, Lin Ma, Weiqiang Zhang, Zhuo Chen, and Yihua Yan. Convolutional neural network for classification of solar radio spectrum. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 198–201. IEEE, 2017.

Zhiyuan Dang, Cheng Deng, Xu Yang, Kun Wei, and Heng Huang. Nearest neighbor matching for deep clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13693–13702, 2021.

Mrinal Das, Suparna Bhattacharya, Chiranjib Bhattacharyya, and Gopinath Kanchi. Subtle topic models and discovering subtly manifested software concerns automatically. In *International Conference on Machine Learning*, pp. 253–261. PMLR, 2013.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html, 2007.

Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pp. 473–483, 1992.

Ryuhei Hamaguchi, Ken Sakurada, and Ryosuke Nakamura. Rare event detection using disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9327–9335, 2019.

Swapnil Hingmire and Sutanu Chakraborti. Sprinkling topics for weakly supervised text classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 55–60, 2014.

Christian Monstein. Catalog of dynamic electromagnetic spectra. *Physics, Astronomy and Electronics Work Bench*, 2011.

Horea Mureșan and Mihai Oltean. Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 10:26–42, 06 2018. doi: 10.2478/ausi-2018-0002.

Sungwon Park, Sungwon Han, Sundong Kim, Danu Kim, Sungkyu Park, Seunghoon Hong, and Meeyoung Cha. Improving unsupervised image clustering with robust learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12278–12287, 2021.

Ian Porteous, David Newman, Alexander Ihler, Arthur Asuncion, Padhraic Smyth, and Max Welling. Fast collapsed gibbs sampling for latent dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 569–577, 2008.

Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Scan: Learning to classify images without labels. In *European Conference on Computer Vision*, pp. 268–285. Springer, 2020.

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.

Stephen M White. Solar radio bursts and space weather. *Asian Journal of Physics*, 16:189–207, 2007.

Yongqin Xian, Christoph H Lampert, Bernt Schiele, and Zeynep Akata. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE transactions on pattern analysis and machine intelligence*, 41(9):2251–2265, 2018.

A Appendix

A.1 Notations Table

Symbol	Description
N, K, V	number of images, number of events, and number of visual words respectively.
H	number of image levels.
α, κ, ζ	hyper-parameters of Dirichlet distributions.
η_i	hyper-parameter of multinomial distribution.
ϱ	hyper-parameter of beta distribution.
ψ_v, π_k	hyper-parameters of bernoulli distribution.
σ, ω	hyper-parameters of dirichlet distribution of size H-1.
θ	$N \times H \times K$ matrix indicating image-event distribution.
β	$K \times V$ matrix indicating event-visual word distribution.
ϕ	binary random vector of event-visual word.
ξ_{ih}	denotes binary random vectors.
δ	threshold.
$-$	Super-script denotes that in all counts the current patch is excluded.
$n_{...kj}^{-ij}$	number of times patch type j is associated with event k.
$n_{...k.}^{-ij}$	number of times event k is used in the whole image dataset.
$n_{i.g_{ij}k.}^{-ij}$	number of times event k and g_{ij} are used.
$n_{i.g_{ij}..}^{-ij}$	number of times g_{ij} is used.
$n_{ihk..}^{-ij}$	number of times event vector indexed by k is used.
$n_{i.h..}^{-ij}$	number of visual words in the image level h.

A.2 Sampling

Sampling of ξ can be defined as follows:

$$\begin{aligned} p(\xi_{ihk} = 1 | z, \xi^{-ihk}) &\propto p(z_i | \xi_{ihk} = 1, \xi^{-ihk}) p(\xi_{ihk} = 1 | \xi^{-ihk}) \\ &\propto \frac{\Gamma(\sum_{s \neq k} \xi_{ihs}\alpha + \alpha)}{\Gamma(\sum_{s \neq k} \xi_{ihs}\alpha + \alpha + n_{i.h..}^{-ij})} \frac{\sum_w \sum_l \xi_{wlk} + 1}{\sum_w \sum_l 1 + 1 + \varrho} \end{aligned} \tag{9}$$

Similarly we can sample ϕ.

Sampling of g: The posterior probability of selecting a event vector for a patch can be defined as follows:

$$\begin{aligned} p(g_{ij} = h | g^{-ij}, z, \xi_{ih}) &\propto p(z_{ij} | g_{ij} = h, z^{-ij}, \xi_{ih}) p(g_{ij} = h | g^{-ij}) \\ &= \frac{\xi_{ihz_{ij}}\alpha + n_{i.hz_{ij}.}^{-ij}}{\sum_k \xi_{ihk}\alpha + n_{i.h..}^{-ij}} p(g_{ij} = h | g^{-ij}) \end{aligned} \tag{10}$$

A.3 Algorithms

We have mentioned our heuristic algorithm color contour detection defined in Algorithm 2. Prediction algorithm for rare events detection in images defined in Algorithm 3 and for multi-class image classification is defined in Algorithm 4. The generative process of synthetic burst data is defined in Algorithm 5.

A.4 Mapping between visual words and patches

We have considered color pixel values as a bag of visual words $\mathcal{V}$ for our Guided MCMC for Bayesian Models. For the sake of simplicity let us assume that a patch j is of size $\mathcal{S}$ x $\mathcal{T}$. Now we need to compute euclidean distance between each pixel value present in patch j and all visual words present in bag of visual words $\mathcal{V}$. Next we would assign visual word index v to patch j based

Algorithm 2 Heuristic algorithm to detect solar bursts in solar radio images sans labels.

```
Input: Area_Threshold, M
Output: Predicted labels
procedure COLOR CONTOUR DETECTION
    Declare Low and High threshold of Red color.
    Initialize Mask= Convert RGB to HSV using low and high threshold of red color.
    for images m = 1 to M do:
        // Find all contours by applying Mask into image m.
        if any contour area ≥ Area_Threshold then:
            Predict current image as Burst
        else
            Predict current image as Non-Burst
        end if
    end for
end procedure
```

Algorithm 3 Prediction algorithm for rare events detection in images using our method.

```
Input: θ, H, r, δ, M
Output: Predicted labels
procedure PREDICT
    for images i = 1 to M do:
        Initialize Pred_label='Non-Rare Class'
        for levels h=1,2,...H do:
            prob_val = θ[i][h][r]
            if prob_val ≥ δ then:
                Update Pred_label='Rare Class'
                break
            end if
        end for
        Print Pred_label
    end for
end procedure
```

Algorithm 4 Prediction algorithm for unsupervised multi-class image classification using our method.

```
Input: θ, H, K, M
Output: Predicted labels
procedure PREDICT
    for images m = 1 to M do:
        Declare prob_val_vector as a zero vector of size 1 x K.
        for levels h=1,2,...H do:
            for events k = 1 to K do:
                prob_val_vector [k]+ = θ[m] [j] [k]
            end for
        end for
        Pred_label_class_index=arg max(prob_val_vector)
        Print: Pred_label_class_index
    end for
end procedure
```

Algorithm 5 Generative process for synthetic solar radio image generation. A palette V represents size of colors. A patch K represents section of image region. We will defined an image size as R number of rows and C no of columns. We will use $alpha$ and $beta$ priors of beta distribution and set them both as 1. We will defined N as number of image generation happened so far. We will defined y as number of Burst images generated so far. Burst Probability(BRYO)=[0.1,0.5,0.2,0.2],Non-Burst Probability(BRYO)=[0.98,0.007,0.006,0.007].We will defined dictionary of colors(color_palette) which contains Blue,Red,Yellow and Orange(BRYO) which will act like a vocabulary.

1: Choose $Ic(Image\_Category) \sim Beta(\alpha\_prior + y, \beta\_prior + N - y)$
2: If Ic is Burst then increment y and do following for all columns in RC image generation:
  1. Initial Burst_prob_success set to 0.3
  2. Choose $Column\_Category \sim Bernouli(Burst\_prob\_success)$
    (a) If $Column\_Category$ is Burst Then
      - Update $Burst\_prob\_success - = 0.2$
      - Do following
        i. Choose $patch \sim Random(Burst\_Palette\_Probab\_Dis)$
        ii. For every twenty rows update $Burst\_Palette\_Probab\_Dis = [B + 0.18, R - 0.1, Y - 0.04, O - 0.04]$
    (b) else
      - Burst_prob_success+=0.1
      - Do following
        i. Choose $patch \sim Random(Non\_Burst\_Palette\_Probab\_Dis)$
        ▷ Note: we set lower bound of Burst_prob_success=0 and above updated probabity will be used for next Column_category sample.
3: If Ic is Non-Burst do following for all columns
  1. Choose $patch \sim Random(Non\_Burst\_Palette\_Probab\_Dis)$
4: update N+=1

on the overall minimum distance of all visual words. We can also add some percentage threshold t which tells if t% of pixel values are closed to visual word index v then assign that particular patch j to v.

A.5 Datasets Description

Solar Radio Burst Image Dataset Description: There are two classes in solar radio spectrograph. 1)Burst class and 2)Non-Burst class. Each RGB image is the size of 1273width x 833height after removing the axis of time and frequency. This dataset contains 50 Burst and 50 Non-Burst images.

Synthetic Burst dataset Description: We have tried to generate synthetic solar radio Burst and Non-Burst images. The detailed generative process can be found in Algorithm 5 in the appendix section. Samples of original as well as synthetic Burst and Non-Burst images can be seen in Figure 4 and Figure 5 in the appendix section A.7. We have generated four synthetic Burst datasets each containing 1000 color images of size 1200width x 800height. 1)Syn_Burst_Data_1 contains 475 Burst and 525 Non-Burst images. 2)Syn_Burst_Data_2 contains 259 Burst and 741 Non-Burst images. 3)Syn_Burst_Data_3 contains 90 Burst and 910 Non-Burst images. 4)Syn_Burst_Data_4 contains 25 Burst and 975 Non-Burst images which mean only 2.5% of images in this dataset belong to Burst class. So that we can evaluate all baselines models as well as our model to measure how they are performing under the degree of high Burst rareness in the synthetic datasets. **Reasons for synthetic data generation:** There are two main reasons for synthetic data generation. The first reason is to challenge our model to check whether it will be able to detect Burst if more noise signals are present in synthetic Burst images, or what happens when there is no Burst at all only noise signals are present and also to check whether our model to detect different shape variation in synthetic Burst images. The second reason is that since we have run our GSTM model only on 100 images because of the unavailability of labeled data.

Fruits Dataset Description: We have also used a simple dataset like the Fruits dataset to check if our model can be applied to perform unsupervised image classification or not. We have created two fruit datasets 1)Fruit_DS_1 has 1947 images and 2)Fruit_DS_2 has 1507 images, using famous fruits360 dataset (Mureșan & Oltean, 2018). In Fruit_DS_1 we have used four fruits category Huckleberry(490 images),Red Apple(492 images),Banana(490 images),Watermelon(475 images) . Each image size is 100x100. In Fruit_DS_2, we have used the same number of Apple, Banana, Watermelon images from Fruit_DS_1 where the only change is that we have made the huckleberry class rare by reducing no of images from 490 to 50.

Aeroplane-Sky Dataset Description: We wanted to test our model on whether an object is present or not in an image for that we have created aeroplane in air dataset. We have randomly selected 25 images of aeroplanes where aeroplane present in the air from pascal-voc 2007 dataset (Everingham et al., 2007) and synthetically generated 475 clear blue sky images of size 100×100.

A.6 Experimental Setup

A.6.1 Experimental Settings for our model

We have used the same hyper-parameters settings for all datasets only the number of events is different. α is a Dirichlet hyper parameter of per image-event distribution is set to 0.1. κ is a Dirichlet hyper parameter of per event-visual word distribution is set to 0.5. For Burst and Synthetic Burst dataset, we have computed the frequency of the values of each pixel in the entire dataset, and based on that we have selected four different pixel values of colors Blue, Red, Yellow, and Orange. In this case, rare visual words are Red, Yellow, and Orange where a non-rare visual word is Blue. The number of events we have set is 10 and the guidance event index is set to 10^{th} index. δ value was set to 0.001 and we have used Algorithm 3 for prediction. In the fruit dataset, there are four fruits categories and we have manually set universal color values for each fruit for example if the fruit category is red apple we have used $[255, 0, 0]$ as a visual word for the red apple category and so on. The number of events we have set is 5. Visual words for this dataset are Red, Yellow, Blue, Green and White pixel values. White was chosen as non-rare visual word because it was present as background color. Now for each fruit category, we have set its own event which means for each event k we have set 1 in ϕ if visal word v belongs to event k and rest all 0 . We have used Algorithm 4 for prediction. For aeroplane dataset, we have considered White, Grey, and Black pixel values as rare visual words and Sky-blue pixel value as non-rare visual word, the number of events we have set is 2 and the guidance event index is set to $2nd$. δ value was set to 0.001 and we have used Algorithm 3 for prediction.

A.6.2 Baselines Experimental Settings

We have used the publicly available implementation of all three baselines algorithms, with default hyper-parameters settings mostly and augmentation_strategy is simclr for all cases. SCAN was used as an off-the-shelf clustering algorithm for RUC. In the original Burst dataset comparison, for SCAN(clustering and self-labeling step we have kept the same batch and augment size), RUC and NNM models we have used batch size as 10, augment size as 96, and confidence threshold as 0.5. In fruit dataset comparison, for SCAN model we have used augment size 96 batch size 11 in clustering step and batch size 137 in self-labeling step, confidence threshold is 0.6, for RUC model batch size is 137 augment size set to 32 confidence threshold is 0.99, For NNM model batch size, is 11 and augment size is 96. In aeroplane dataset case, for SCAN, NNM models we have used batch size as 50, augment size as 96, and confidence threshold as 0.99 in SCAN, and RUC model we have used batch as 25 and confidence threshold as 0.75. In the Syn_Burst_Data_1, for SCAN model we have used batch size is 10, augment size is 96 and confidence threshold set to 0.5. for RUC model we have used batch size as 10 and confidence threshold as 0.5 and augment size as 70. for NNM model we have used batch size as 10 and augment size as 70. In the Syn_Burst_Data_2, for SCAN model we have used batch size is 100 in clustering step and in self labeling step batch size is set to 250, augment size is 70 and confidence threshold set to 0.6. for RUC model we have used batch size as 10 and confidence threshold as 0.6 and augment size as 70. for NNM model we have used batch size as 10 and augment size as 70. In the Syn_Burst_Data_3, for SCAN model we have used batch size is 100 in clustering step and in self labeling step batch size is set to 250, augment size is 70 and confidence threshold set to 0.75. for RUC model we have used batch size as 10 and confidence threshold as 0.75 and augment size as 70. for NNM model we have used batch size as 10 and augment size as 70. In the Syn_Burst_Data_4, for SCAN model we have used batch size is 10 in

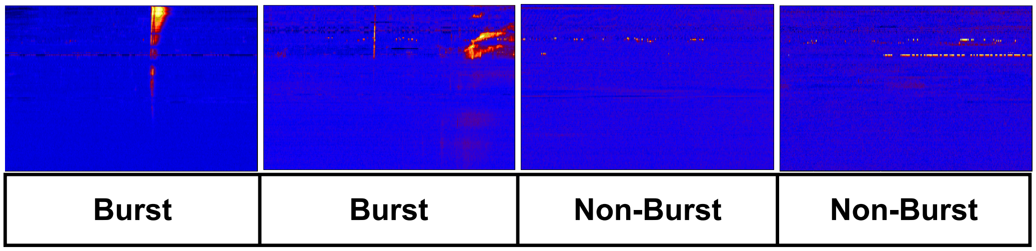

Figure 4: Samples of four solar radio images are shown here. The first two images contains Burst and last two images are Non-Burst which contains only noise signals.

clustering step and in self labeling step batch size is set to 10, augment size is 96 and confidence threshold set to 0.75. for RUC model we have used batch size as 10 and confidence threshold as 0.5 and augment size as 70. for NNM model we have used batch size as 10 and augment size as 70.

A.7 VISUALS

The events generated by Guided MCMC for Sparse Bayesian Models can be seen in Table 5. We have shown samples for original burst dataset in Figure 4 and samples for synthetic burst dataset can be seen in Figure 5. The event-visual words ϕ used as guidance on synthetic burst, aeroplane-sky, and fruit datasets are shown in Table 6, 7, 8 respectively.

Table 5: Events Generated by Guided MCMC for Sparse Bayesian Models in which top visual words are shown as background colors with their corresponding probability values. a) Events generated on synthetic burst dataset, here we have shown b) Events generated on aeroplane-sky dataset, c) Events generated on fruit dataset.

Non-Burst Event	Burst Event
0.3492	0.1966
0.2895	0.1801
0.2146	0.1562
0.1377	0.1391
0.0011	0.0555
0.0010	0.0486
0.0009	0.0448
0.0008	0.0383
0.0008	0.0377
0.0007	0.0372

a)

Sky Event	Aeroplane Event
1	0.7807
0	0.1954
0	0.0238
0	0

b)

Huckleberry Event	Apple Red Event	Banana Event	Watermelon Event	Other Event
1	1	1	1	1
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

c)

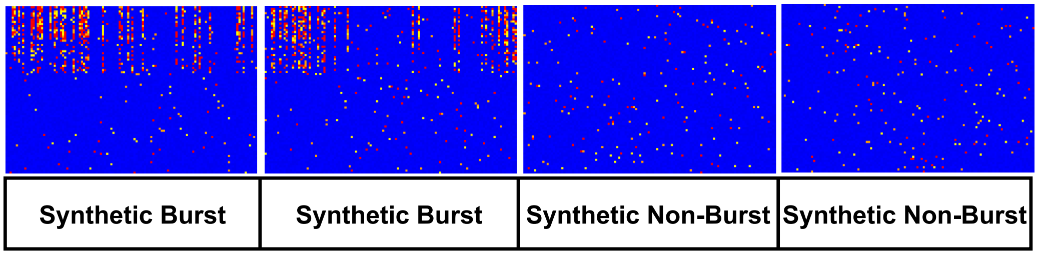

Figure 5: Samples of four synthetic solar radio images are shown here. The first two images contains Burst and last two images are Non-Burst which contains only noise signals.

Table 6: Events generated on synthetic Burst dataset by providing a) event-visual ϕ guidance to our model which will generate b) event-visual words distribution. In the guidance part we have set index 0 as Non-Burst event and index 1 as Burst-event. For Non-Burst event we have filled 1 in all indexes and for Burst event, we have filled 1 only in place of rare visual words indexes i.e red, yellow, and orange. Notice that the index 1 in guidance of event-visual ϕ and event-visual words distribution corresponds to Burst event. Now using index 1 in image-event distribution can help to detect bursts events in solar radio images.

Visual Words	a) Guidance: Index 0 Non-Burst Event	a) Guidance: Index 1 Burst Event	b) Event-Visual words distribution: Index 0 Non-Burst Event	b) Event-Visual words distribution: Index 1 Burst Event
1	1	0	0.3492	0
2	1	0	0.2895	0
3	1	0	0.2146	0
4	1	0	0.1377	0
5	1	1	0.0006	0.1966
6	1	1	0.0005	0.1801
7	1	1	0.0010	0.1562
8	1	1	0.0011	0.1391
9	1	1	0.0008	0.0555
10	1	1	0.0008	0.0486
11	1	1	0.0009	0.0448
12	1	1	0.0006	0.0383
13	1	1	0.0002	0.0377
14	1	1	0.0004	0.0372
15	1	1	0.0005	0.0347
16	1	1	0.0007	0.0305

a) Guidance

b) Event-Visual words distribution

Table 7: Events generated on aeroplane-sky dataset by providing a) event-visual ϕ guidance to our model which will generate b) event-visual words distribution. In the guidance part we have set index 0 as Sky event and index 1 as aeroplane-event. For sky-event we have filled 1 in non-rare visual word index i.e. sky blue color and for aeroplane event, we have filled 1 in all rare visual words indexes i.e white, black, and grey. Notice that the index 1 in guidance of event-visual ϕ and event-visual words distribution corresponds to aeroplane event. Now using index 1 in image-event distribution can help to identify aeroplane is present in images or not.

Visual Words	a) Guidance: Index 0 Sky Event	a) Guidance: Index 1 Aeroplane Event	b) Event-Visual words distribution: Index 0 Sky Event	b) Event-Visual words distribution: Index 1 Aeroplane Event
1	1	0	1	0
2	0	1	0	0.0238
3	0	1	0	0.1954
4	0	1	0	0.7807

Table 8: Events generated on fruit dataset by providing a) event-visual ϕ guidance to our model which will generate b) event-visual words distribution. In the guidance part we have set fruit category for each index. for example in index 0 we have set 1 to blue as rare visual word for huckleberry class and rest we have filled 0 in all indexes of huckleberry class. Similarly for other fruits we did the same procedure. Notice that the each index in guidance of event-visual ϕ and event-visual words distribution corresponds to same fruit event. Now using image-event distribution we can identify which image belongs to a particular fruit class.

Visual Words	a) Guidance: Index 0 Huckleberry Event	a) Guidance: Index 1 Apple Red Event	a) Guidance: Index 2 Banana Event	a) Guidance: Index 3 Watermelon Event	b) Event-Visual words distribution: Index 0 Huckleberry Event	b) Event-Visual words distribution: Index 1 Apple Red Event	b) Event-Visual words distribution: Index 2 Banana Event	b) Event-Visual words distribution: Index 3 Watermelon Event
1	1	0	0	0	1	0	0	0
2	0	1	0	0	0	1	0	0
3	0	0	1	0	0	0	1	0
4	0	0	0	1	0	0	0	1