Burnt Forest Estimation from Sentinel-2 Imagery of Australia using Unsupervised Deep Learning

Nosheen Abid*†‡, Muhammad Imran Malik†‡, Muhammad Shahzad†‡‖, Faisal Shafait†‡, Haider Ali§, Muhammad Mohsin Ghaffar¶, Christian Weis¶, Norbert Wehn¶, and Marcus Liwicki*

*EISLAB Machine Learning, Luleå Tekniska Universitet, Sweden

‡Deep Learning Lab, National Center of Artificial Intelligence, National University of Sciences and Technology, Pakistan

†School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan

‖Department of Aerospace and Geodesy, Technical University of Munich (TUM), Munich, Germany

¶Department of Electrical and Computer Engineering, TU Kaiserslautern, Germany

§Department of Computer Science, Johns Hopkins University, USA

***Abstract*—Massive wildfires not only in Australia, but also worldwide are burning millions of hectares of forests and green land affecting the social, ecological, and economical situation. Widely used indices-based threshold methods like Normalized Burned Ratio (NBR) require a huge amount of data pre-processing and are specific to the data capturing source. State-of-the-art deep learning models, on the other hand, are supervised and require domain experts knowledge for labeling the data in huge quantity. These limitations make the existing models difficult to be adaptable to new variations in the data and capturing sources. In this work, we have proposed an unsupervised deep learning based architecture to map the burnt regions of forests by learning features progressively. The model considers small patches of satellite imagery and classifies them into burnt and not burnt. These small patches are concatenated into binary masks to segment out the burnt region of the forests. The proposed system is composed of two modules: 1) a state-of-the-art deep learning architecture for feature extraction and 2) a clustering algorithm for the generation of pseudo labels to train the deep learning architecture. The proposed method is capable of learning the features progressively in an unsupervised fashion from the data with pseudo labels, reducing the exhausting efforts of data labeling that requires expert knowledge. We have used the real-time data of Sentinel-2 for training the model and mapping the burnt regions. The obtained F1-Score of 0.87 demonstrates the effectiveness of the proposed model.**

***Keywords*—Unsupervised; Deep Learning; Australia; Forest Fire; Wildfire; Sentinel-2; Aerial Imagery**

I. Introduction

Australia is, more than any other, a fire continent [1]. It has faced an annihilating beginning of a gigantic fire within the last quarter of 2019, which burnt over 5.8 million hectares of forests, mostly in Victoria(VIC) and New South Wales (NSW). In general, the number of fire alerts in Australia has increased in the past two decades due to an increase in humidity, drought, record heat, and high winds [2]. Similar to Australia, forests in other continents have historically burned up to approximately 5% in the previous decade [3] essentially devastating biodiversity, timberland riches, and human settlements [4].

Considering the severity of the circumstances and disadvantages of the furious blaze [5], the research community has actively worked on the issue. Many methods and solutions have been designed for detecting and monitoring the woodland fire by different sources like Radio Detection and Ranging (RADAR), Light Detection and Ranging (LiDAR), and optical imagery [6]. The primary sources of remote optical imagery are Unmanned Aerial Vehicles (UAV) and satellites. Satellite imagery, captured through multispectral sensors, is particularly used worldwide for forest fire detection and assessment.

Satellite imagery-based burnt area classification algorithms can be generally divided into two major categories; rule-based methods and machine learning methods. Rule-based methods are a combination of the spectral response of burnt region mostly in short wave infrared (SWIR) and near-infrared (NIR) bands of the satellite imagery. This is because fire has a significant reflectance in the SWIR and NIR bands. The spectral responses in these bands are pretty helpful in detecting sound vegetation and burned regions. In burning areas, a significant drop in values is observed in NIR reflectance and a rise in SWIR reflectance after burning. This response is because of the sensitivity of the NIR band to chlorophyll substance of healthy plants, and SWIR captures the moisture of soil and vegetation [7]. These multispectral bands, along with visual bands (red, green, and blue), are commonly used in the indices applied to detect burnt regions in satellite imagery. The most commonly used indices for the purpose are Normalized Burned Ratio (NBR) [8], the Mid-InfraRed Burn Index (MIRBI) [9], and the Modified Burned Area Index (BAIM) [10]. For detection of burnt areas from an aerial view, mostly pre-event imagery of the scene is used along with the post-event to detect the changes with the help of empirically calculated thresholds. Here, an insufficient choice of a pre-event scene may lead to misclassifications. Additionally, these traditional rule-based approaches are sensitive to noise, like cloud cover, and require exhaustive preprocessing of a massive corpus of data–thereby making the task more challenging.

In the recent past, several machine learning-based techniques have been designed to map burnt regions using remotely sensed imagery. The lately designed algorithms MCD64AI at 500 m resolution [11] and FIRECCI51 at 250-meter resolution [12] create temporal composites for capturing the lasting changes, sift low-quality pixels, and combine these processed pixels with active blaze identified using the MODerate resolution Imaging Spectroradiometer sensor (MODIS).

For FireCCI51, initially, some candidate pixels for the burnt area are detected. Later the neighboring burnt pixels of the candidate pixels are identified using a pixel growing algorithm. For MCD64A1C6, steps in series are followed, including the region growing procedure. The use of the region growing technique is a very common practice in traditional approaches for mapping burnt regions [13], [14], [15]. Many other algorithms have been presented and used at regional levels. For instance, One-Class Support Vector Machine [16] is used for reducing the omission error produced by the omission of the active blaze. It has minimized the requirement of using the region growing technique to make the approach comparatively easier. However, the proposed method used temporal composite for avoiding cloud cover and cloud shadows leading to discarding some information that could be useful. Furthermore, sensors vary from each other in characteristics. Most of the traditional approaches are sensitive to the particular sensors they are designed and refined for, and adaptation of these algorithms to different sensors becomes a challenging task. Despite improvements over the years in the algorithms for burnt area mapping, there are still some facets that need improvements and/or are outside the limitations of the traditional methods. Specifically, the burnt zone mapping tools would be more useful for larger and steady time-series data, better uncertainty estimation, and mapping blaze areas and combustion completeness [17]. It raises the need to have a method that is scalable as well as adaptable to variations. Deep learning (DL) is capable of addressing the above-stated limitations [18].

Today Deep Learning (DL) techniques are rapidly becoming state-of-the art for learning variant and complex features across various domains [19]. Computer-vision problems of object detection, localization, and recognition are thrived by Convolutional Neural Networks (CNN) [20]. In the domain of remote sensing, CNNs are used in emerging applications of land-cover classification [21], segmentation of buildings, roads [22] and small objects [23], reconstruction of missing information in the data [24], cloud-cover detection [25] and cloud shadows and effective utilization of Spatio-temporal satellite data [26], [27], [28], [29]. On a similar note, burnt land mapping and dating have also been addressed by using deep learning [30]. Pinto et al. [31] combined CNNs and Long-Short-Term-Memory (LSTM) with U-Net based architecture for mapping and dating burnt regions using multispectral imagery. Similarly, Knopp et al. [32] have segmented burnt land from mono-temporal Sentinel-2 using U-Net based architecture.

Even though the deep learning methods are becoming state-of-the-art, they carry a few limitations. 1) They are generally supervised and require a huge amount of labeled data for training the model. Data labeling is time-consuming and an exhaustive task. Furthermore, in many instances, it requires expert knowledge, which is hardly available at scale. 2) They are domain-specific. Their performance diminishes radically when applied to a diverse dataset of the same problem. To bargain with these issues, the concept of Curriculum Learning [33] has been used by a few researchers [34]. In this paper, we have also used a similar concept. We performed burnt forest estimation by combining the state-of-the-art machine learning and computer vision methods with the concept of CL using Sentinel-2 Imagery of Australia. We have performed unsupervised patch-based classification of burnt and unburnt patches in satellite imagery. The process itself covers three stages, including the selection of training examples, computing the discriminative features, and classify the burnt and unburnt regions.

The proposed unsupervised method is composed of a deep learning architecture, a clustering algorithm, and a selection operation based upon curriculum learning for burnt region classification. The model takes satellite image patches as input and classifies them into "burnt" and "not burnt" class. It is assumed that the input patch is not labeled. Therefore, a clustering algorithm is used to tackle the issue of labeling. Firstly, pre-trained deep learning architecture is used to extract feature vectors of patches. Secondly, these feature vectors are clustered into two categories using state-of-the-art clustering techniques to generate pseudo-labels, assuming them to be burnt and not burnt. The pseudo-labels are used as a new identity of the patch. Initially, the pre-trained deep learning model is trained on the ImageNet dataset, which does not contain aerial imagery. Hence, the pre-trained model may not extract good feature vectors resembling burnt and not burnt region of aerial imagery, resulting in loose clusters in the feature space. Our hypothesis is that clustering will give better distribution of burnt and non-burnt patches than random division. The main idea of the approach is to iteratively improve the feature extractor by fine-tuning the deep learning model on representative samples from each cluster. A selection operation is used for selecting the samples from generated clusters for fine-tuning. The selection operation selects the samples present near the centroids of the clusters. These samples indicate the prominent features of the respective cluster. Fine-tuning of the model with selected samples and respective pseudo-labels makes the model learn the discriminative features between the two clusters. As a result, better discrimination is achieved between burnt and not burnt regions. When the model converges, one of the clusters will belong to the "burnt" and the other to the "not burnt" class. The major contributions of our work include:

1) A progressive deep learning model for burnt region classification from multispectral aerial imagery.
2) An unsupervised architecture removes the need for data labeling, which is a primary requirement of state-of-the-art supervised deep learning based methods.
3) A patch-based Sentinel-2 imagery dataset of burnt Australian regions from the 2020 fire incident has been developed and will be publicly available.

II. Materials and Methods

A. Study Area

In this study, different regions of Victoria (VIC) and New South Wales (NSW) have been considered for analyzing the wild bushfire. Figure 1 graphically shows the regions selected for this analysis while Table I lists down the precise geographical locations in the standard latitude/longitude (WGS84) coordinates and surrounding cities. The NSW and Victoria regions are chosen as these were the worst-hit states of Australia affected by the massive fire of 2019-2020. It has burnt more than 5 million hectares of bush, forests, and parks across the state and destroyed more than 2,000 houses. The wildfire was mostly located at the coast of the Tasman Sea in NSW. The windy conditions and hot weather added fuel to the fire and

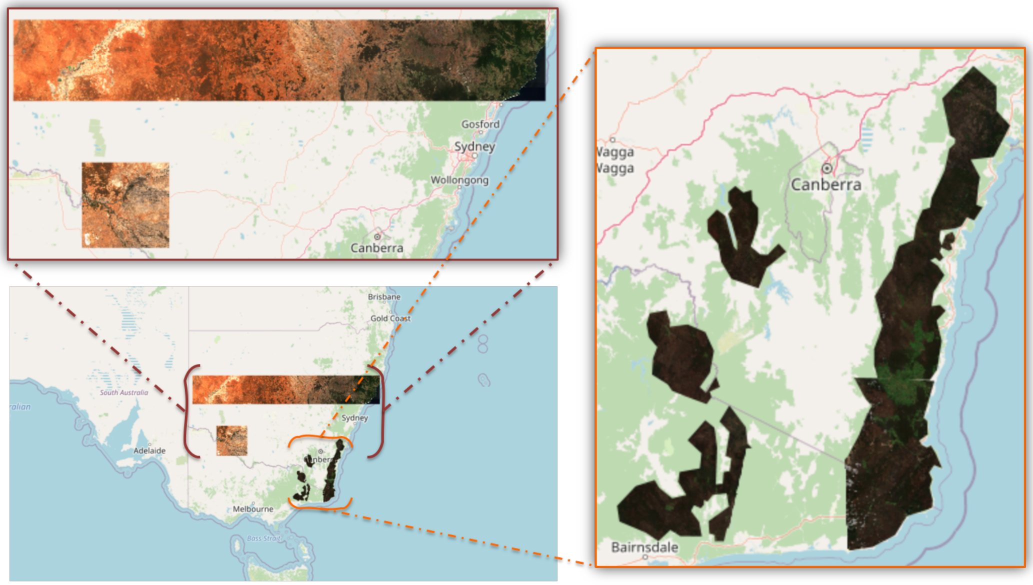

Fig. 1: An illustration of the regions of New South Wales (NSW) and Victoria is considered in this study for analyzing and training the deep learning model. The rectangular regions indicate the area of early 2019 and are considered as not burnt. Whereas the other polygons in NSW and Victoria are of the 1st quarter of 2020 and considered as burnt regions as a result of the massive fire.

Class	**Region**	**Latitude**	**Longitude**	Cities			
				East	**West**	**North**	**South**
Not Burnt 2019	Mymagee	-32.14	146.72	Crowdy Head	Broken Hill	Gowang	Mount Hope
	Balrang	-34.70	143.64	Maude	Rabinvale	Corrong	Winlatin
Burnt 2020	-	-35.76	148.43	Yarran-gobily	Buddong	-	Cabramurra
	Torn Groggin	-36.53	1447.92	Murray Gorge	-	Nariel Vally	-
	East-South Coast NSW	-36.40	149.59	Narooma	-	Bundanoon	Tamboon

TABLE I: The table shows the details of the considered NSW regions of 2019 and 2020. Two polygons are considered from 2019 for "Not Burnt" class, whereas three polygons are considered from 2020 for "Burnt" class. These regions are used for training and testing the unsupervised deep learning model.

resulted in an uncontrollable situation. The massive bushfire raged the area, including the Australian capital Canberra, for weeks and months. In Victoria, the fire affected 1.2 million hectares by early January 2020 [35]. The generated smoke had drastically polluted the environment, air quality, and satellite imagery. According to Swiss-based group AirVisual [36], the quality of the polluted air in Canberra (capital) was rated as the 3rd worst of all major global cities on January 3, 2020. The satellite images from early January 2020 manifested significant dissemination of smoke from firestorms in NSW and Victoria and spread far away to New Zealand as reported in the BBC report [37].

B. Data and Pre-Processing

The Sentinel-2 multispectral imagery of the selected regions of NSW and Victoria is used to generate the required dataset. We have visualized the Sentinel-2 imagery and accordingly labeled the unaffected and affected regions from fire. The rectangular regions, as shown in Figure 1, belong to the 1st quarter of 2019 and are considered as unaffected from the wildfire. Whereas the randomly shaped polygon corresponds to the 1st quarter of 2020 and is the affected burnt regions. We have extracted the respective bundle of images considering only equal to or less than 1% of the clouds. From both bundles, the two respective median images are computed. The 2019 median image of the rectangular region is used as an unaffected class, whereas the 2020 median image of irregular polygon regions is used as the burnt forest class. These two images are divided into small patches of size 64x64. We have considered the 12 bands of Sentinel-2, that are, Band 1 – Coastal aerosol, Band 2 – Blue, Band 3 – Green, Band 4 – Red, Band 5 to 7 – Vegetation red edge, Band 8 – NIR, Band 8A – Narrow NIR, Band 9 – Water vapor, Band 11 – SWIR, Band 12 – SWIR (i.e., all the bands are used except the Band – SWIR – Cirrus – 10 as it does not provide the surface information).

These bands are concatenated to make them suitable to feed as input to the deep learning network. The four of these bands are concatenated together to form one channel of size

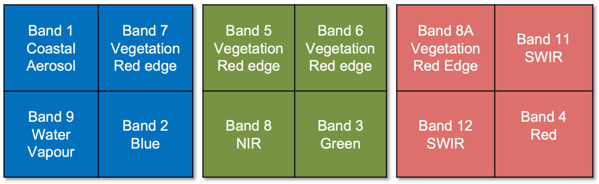

Fig. 2: Shows the considered 12 bands of multispectral satellite imagery of Sentinel-2 into three-channel input. Each channel contains four bands, 1 in each quarter. Each of the red, green, and blue bands is kept in each channel, considering the input configuration of the CNN model. All the patches are preprocessed in this way to make them suitable for the input of the deep learning model.

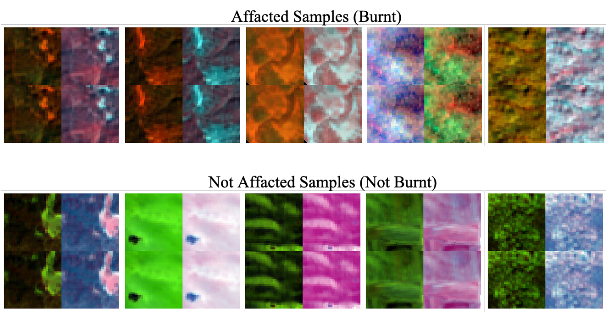

Fig. 3: Shows the visuals of 10 samples of burnt and not burnt from the dataset where each sample is composed of 12 bands of Sentinel-2 imagery concatenated into three channels.

(128x128). Similarly, three-channel sample size (128x128x3) is created out of the chosen twelve bands. Figure 2 graphically shows the bands concatenation. Among the generated data, we have randomly selected 12,000 samples for training and evaluating the employed unsupervised deep learning model. The considered dataset of 12,000 samples includes 6,000 affected (i.e., burnt forest) and 6,000 not affected regions' samples, out of which 8,000 were used for training and 4,000 for testing. Some affected and unaffected samples from the generated dataset are shown in Figure 3.

C. Deep Unsupervised Burnt Forest Learning Scheme

The proposed methodology essentially frames the burnt forest monitoring process in an unsupervised learning manner. It does so by adopting a two-phase procedure: In the first phase, a base deep convolutional neural network (CNN) is patch-wise trained (as an initializer for the subsequent phase) on the relevant dataset to learn robust and distinctive burnt forest features. Subsequently, in the second phase, these features are then input to an unsupervised clustering scheme to perform the grouping of image patches that share similar appearance characteristics. The fundamental underlying idea is to *iteratively* fine-tune this whole feature extraction and clustering scheme in an unsupervised way. The idea has been adopted from computer vision (e.g., [38] [39] [40]) which combines the strengths of transfer learning and latent space representation to enable cross-domain adaptation. The clustering results are treated as *pseudo* labels and are fed back to the network to further fine-tune the base model. The process then continues with increasingly growing training samples with pseudo labels until convergence. Following are the individual steps outlined in a sequential manner:

1) Perform feature extraction using a pre-trained base model to extract robust burnt forest feature representations;
2) Feed the extracted feature representations to an unsupervised clustering to cluster burnt forests from the rest image patches;
3) Refine the obtained clusters to probabilistically retain the representative image patches;
4) The cluster IDs are used to assign pseudo labels to the unlabeled refined image patches;
5) Retrain (i.e., fine-tune) the deep learning module with each refined image patch of every cluster;
6) Extract features of the whole unlabeled training corpus using the fine-tuned model obtained from Step 5.
7) Repeat Step 2 to 6 until the deep learning model is converged.

For Step 1, the adopted base model is the VGG16 model pre-trained on a large-scale ImageNet dataset which is employed for extracting features from the training input image patches. The output of the last convolutional layer is extracted to get feature maps of each sample in the dataset. The extracted features are flattened to get the feature vectors. To cluster these feature representations, a well-known unsupervised k-means clustering algorithm is adopted. The input layer of the VGG-16 is adapted according to remote sensing image patch size and the output layer to the number of clusters that are generated. If we suppose that the features extracted from the training image patches $\{x_i\}_{i=1}^N$ are represented by $\{f_i\}_{i=1}^N$, then in Step 2, these features are clustered using k-means objective function: $\{y_i\}_{i=1}^N \leftarrow \min \sum_{i=1}^N \sum_{k=1}^2 |f_i - c_k|$ where each feature vector is assigned a cluster label $\{y_i\}_{i=1}^N$ on the basis of its minimum distance from the particular centroid c_k, where c is the centroid of the kth cluster. In the current scenario, we have set the value of k to be two so that all the image patches are clustered into two groups, namely burnt forest and the other category. In Step 3 and 4, the obtained clustering IDs are used to assign the pseudo labels that are later used for fine-tuning the CNN model.

Since the employed VGG16 model use model weights that have been trained on a completely irrelevant (natural) dataset, therefore the obtained clusters are quite noisy and cannot be directly used to fine-tune remote sensing images to recognize burnt forest image patches. To cope with this issue, the obtained clusters are passed through a filtering mechanism to prune individual clusters. For this purpose, only those features are retained whose distance to the centroid of the pixel is less than a certain threshold. The refinement process keeps only the feature points near the cluster centroids and thus restricts the CNN to learn only the prominent features and avoid unnecessary noisiness. The image patches belonging to the refined clusters are then subsequently used to retrain the whole network in an unsupervised manner, i.e., with the cluster IDs as pseudo labels, in Step 5. In the next iteration, the updated (fine-tuned) model is used to extract the features from the image patches. With every iteration, the model learns the image features using the pseudo labels of clusters resulting

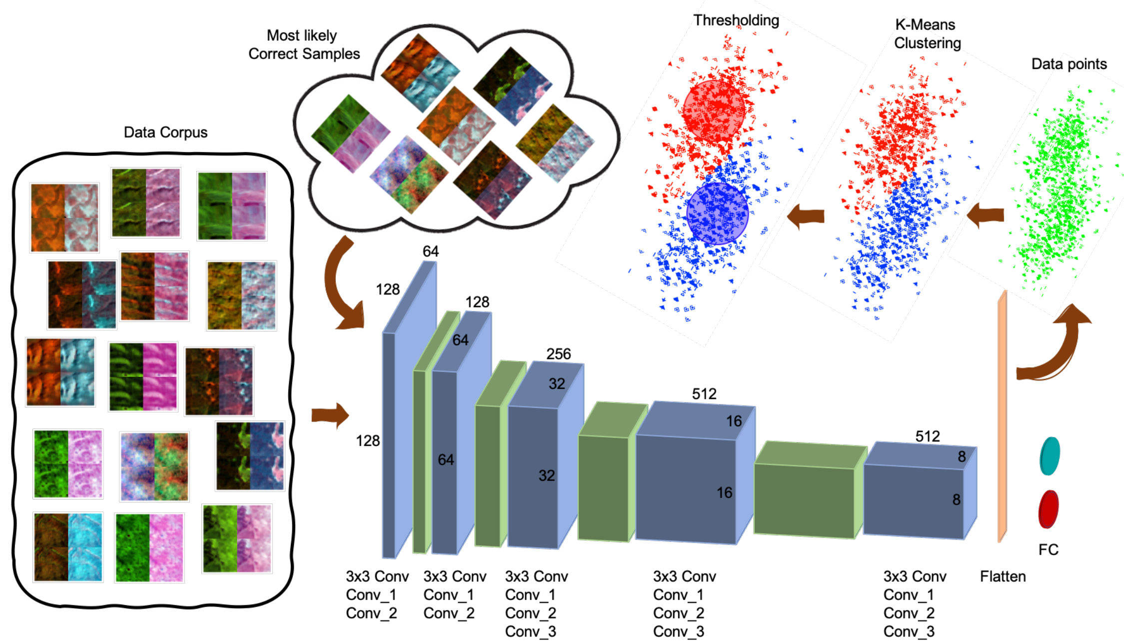

Fig. 4: Shows the three prominent steps of the technique; 1) Deep learning model (VGG-16) to learn and extract the features, 2) Clustering to generate pseudo labels, and 3) Thresholding the samples present near the centroids of the clusters and declaring them as reliable samples.

in comparatively better clusters than the previous iteration. The process thus iterates until the loss of the deep model is converged. See Figure 4 for model visualization.

III. Results and Discussion

A. *Clustering*

Initially, the clusters are generated out of features extracted from a pre-trained deep learning model. The model is trained on an irrelevant domain (ImageNet) which does not contain the satellite imagery, more specifically, burnt forests in Sentinel-2 Imagery. As a result, the clusters are not compact and loosely packed for our input of Sentinel-2 imagery. To evaluate the compactness of the clusters, purity and the Silhouette Score are computed (see Figure 5). Purity is a supervised measure that calculates the ratio of correctly classified samples to the total number of samples for all clusters. Silhouette Score is an unsupervised measure that calculates the ratio on the basis of the distance between each sample within-cluster and the neighboring clusters.

It can be seen in the graph that, in the beginning, the Silhouette Score is a small number, which is 0.07, indicating the lack of compactness in the clusters. With every iteration of fine-tuning of the model, the compactness in the clusters increases (at max to 0.64 at the 7th iteration). After a few iterations, the compactness is saturated. Whereas purity remains consistent throughout between the interval (0.75 - 0.85), indicating the ratio of correctly classified samples. Though the purity stays around 80%, but saturation increases with fine-tuning leading to better cluster segregation.

For further analysis, the Sum of Squared Error (SSE) is calculated, which is also the objective function of K-Means

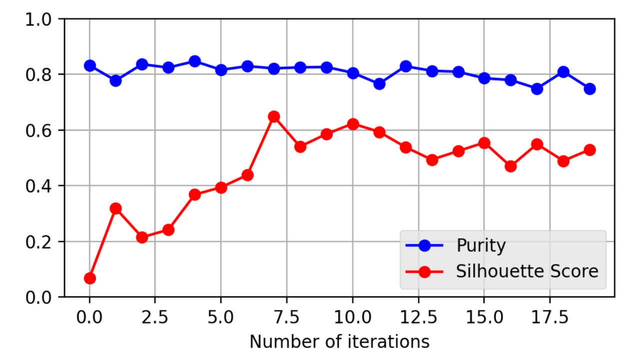

Fig. 5: Graph showing the purity and Silhouette Score of generated clusters for burnt forest and other regions over every iteration of fine-tuning of VGG16.

clustering. It can be seen in Figure 6. Initially, the value of SSE was quite high when clustering was done using pre-trained VGG16. As soon as the model is fine-tuned on a few images of Sentinel-2, the SSE decreased significantly by the one-degree exponent. After that it remains consistent at mean $3.9\text{x}e^{8}$.

As the clusters are loosely packed in the beginning, it is better to use those samples present near the centroids of the clusters for fine-tuning of VGG16. It restricts the model from learning random features and ignores the noisy samples from the clusters. To do so, the dot product is used to find the similarity of every sample within a cluster with its respective centroid. Its value ranges from 0 to 1. If the dot product is greater than or equal to a pre-defined threshold, it is counted as

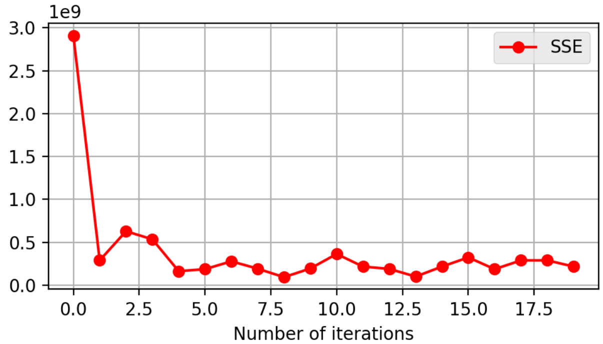

Fig. 6: Graph showing Sum of Squared Error of generated clusters for burnt forest and other regions over every iteration of fine-tuning of VGG16.

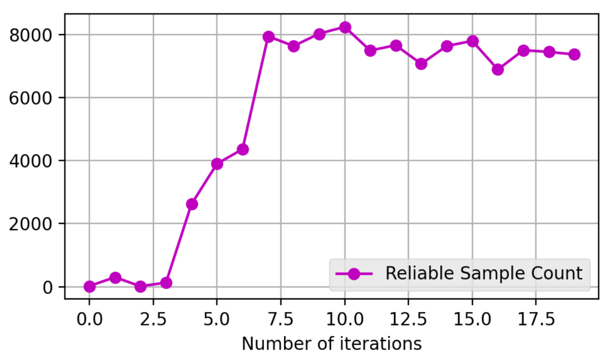

Fig. 7: Graph showing the count of reliable samples over every iteration of fine-tuning of VGG16.

the sample present near the centroid and declared as a reliable sample.

It can be seen in Figure 7 that at the start, only a few, i.e., 22 reliable samples, are extracted using the pre-trained VGG16. It indicates that the clusters are loosely packed. As the VGG-16 gets fine-tuned iteratively, the count of a reliable sample grows. The growth of reliable samples gets saturated after some iterations. After the 7th iteration, the graph remains quite consistent, with a count close to 8,000. It indicates that the clusters contain the majority of the samples from the training corpus of 9,000 samples as a reliable set and declaring only a small fraction of about 1,000 as the noisy ones.

Considering the purity, Silhouette Score and SSE measures, and count of reliable samples, it can be seen that the clusters generated at the 7th and 10th iteration are the best ones. We have fine-tuned the model for 20 interactions. After this, no more improvement in the clustering and fine-tuning is observed.

B. Fine-tuning VGG16

VGG16 is a supervised deep learning architecture that requires labels along with images to train the model. The pseudo labels of generated clusters are used to train VGG16. It can be seen in Figure 8 that the model is reporting almost a very small cross-validation loss on every iteration of fine-tuning VGG-16, considering the pseudo labels as the labels of the patches. It shows that the model is effectively learning

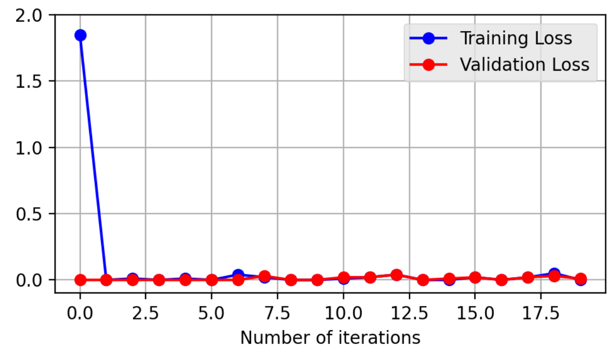

Fig. 8: Graph showing the training and cross-validation loss with pseudo labels over every iteration of fine-tuning of VGG16.

the generated labels of the patches. The model is fine-tuned end-to-end till the classification layer. This fine-tuned model is used for feature extraction in the next iteration, where the last max-pooling layer's output is used to generate the feature vectors for the clustering step.

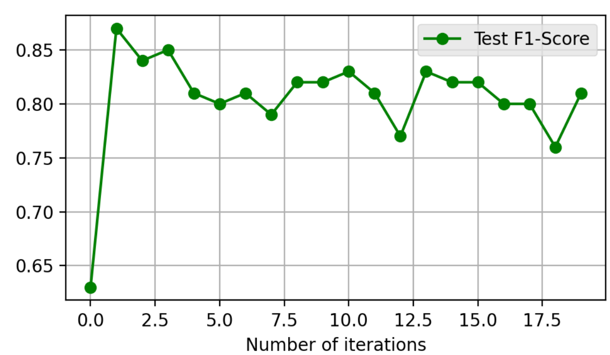

Fig. 9: Graph showing F1-Score on the test corpus over every iteration of fine-tuning of VGG16.

The fine-tuned models generated over 20 iterations are evaluated on a real-time dataset. Figure 9 shows the calculated F1-Score. It can be seen that the model at 1st, 2nd, 3rd, 10th, and 13th iteration gave the top 5 F1-Scores. Considering the top 5 results on test corpus over the 20 iterations, a mean of precision, recall, F1-Score, and accuracy are reported in Table II.

C. Analysis on Sentinel-2 Imagery

We have considered the median Sentinel-2 imagery of three months (Feb 2020 - Apr 2020) for the region, Australian Capital Territory, and South of it, see Figure 10-(a). This area is the worst affected region by the massive wildfire. The top 5 iterations of fine-tuning the model reporting the highest performance on test corpus were deployed to analyze their performance on the considered region. The results can

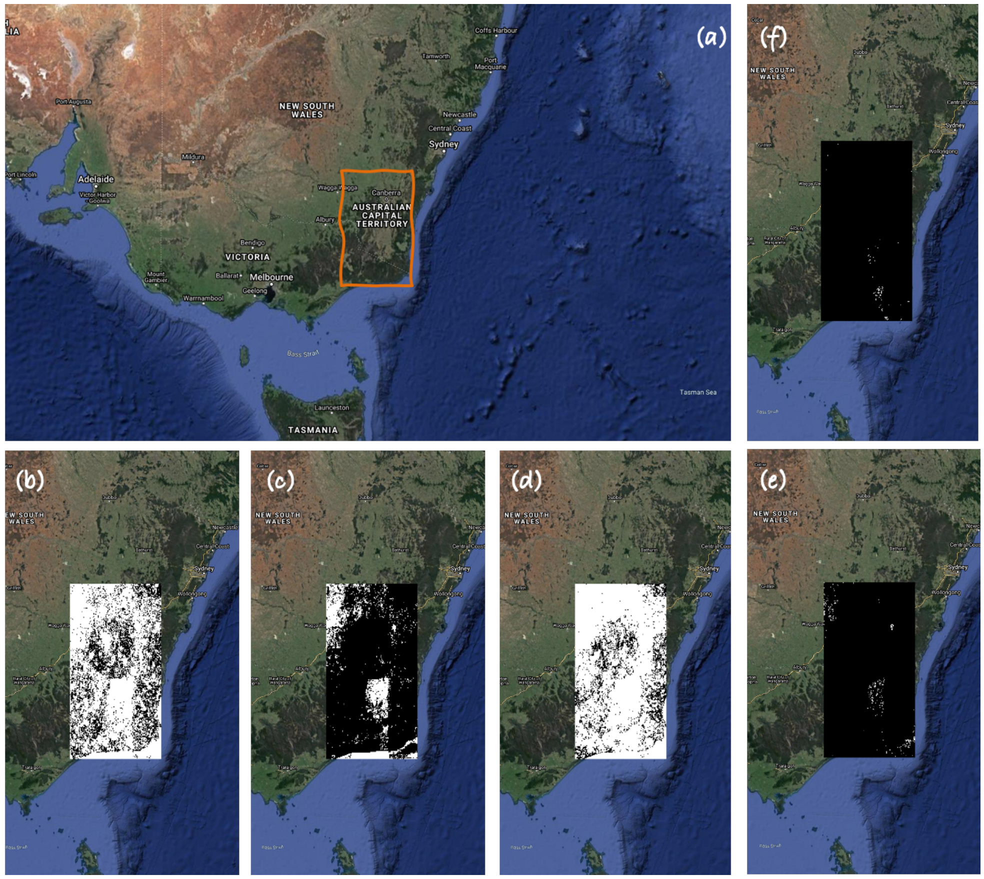

Fig. 10: **(a)** Shows the region considered, covering Australian Capital Territory and South of it in an orange polygon, for testing the fine-tuned models on Sentinel-2 median image of three months (Feb 2020 - Apr 2020). **(b)** Shows the prediction results for the fine-tune iteration of the deep learning model, reporting the highest accuracy on test corpus. **(c)** Shows the prediction results for iteration reporting the 2nd highest accuracy on test corpus. **(d)** Shows the prediction results for iteration reporting the 3rd highest accuracy on test corpus.. **(e)** Shows the prediction results for iteration reporting the 4th highest accuracy on test corpus. **(f)** Shows the prediction results for iteration reporting the 5th highest accuracy on test corpus.

	Precision	**Recall**	**F1-Score**
Class 0	0.86	0.83	0.84
Class 1	0.83	0.86	0.85
Accuracy	-	-	0.85
Macro Avg	0.85	0.85	0.85
Weighted Avg	0.85	0.85	0.85

TABLE II: The table shows the average precision, recall, F1-Score and accuracy for best 5 iterations on the test corpus.

be seen in Figure 10-(b-f). Image (b) of the figure shows the result of 1st fine-tune iteration, reporting the highest accuracy of 0.87 on the test corpus. Similarly, (c) shows the result of the 3rd iteration reporting the 0.85 accuracy, (d) shows the result of 2nd iteration reporting 0.84 accuracy, and (e-f) results of 10th and 13th iterations, each reporting 0.83 accuracy. The accuracy on the test corpus for 1st, 2nd, and 3rd is comparatively higher, but the generated clusters are loosely packed compared to later iterations of fine-tuning, and the count of reliable samples is less than 50 (see subsection III-A). Considering the compactness and good count of reliable samples, the models at iteration 10 and 13 generate better clusters but performance decreases on the test set. This might be because of multiple iterations of fine-tuning, leading the models towards overfitting.

IV. Conclusions

In this paper, we have proposed an unsupervised deep learning technique for mapping the burnt regions of Australia. The method is capable of learning the features progressively from the data without expert knowledge. The proposed solution provides the advantages of supervised deep learning models along with removing the tedious step of data labeling. We are able to achieve the F1-Score of 0.85 with the progressive learning behavior of the model in an unsupervised manner. The real-time Sentinel-2 Imagery is used for training the deep learning architecture and mapping the burnt region of Australia. The method can be applied without any modifications to estimate the burnt forest region in any other region of the world due to its unsupervised nature.

References

[1] S. J. Pyne, *World fire: the culture of fire on earth*. University of Washington press, 1997.

[2] D. Bowman, G. Williamson, M. Yebra, J. Lizundia-Loiola, M. L. Pettinari, S. Shah, R. Bradstock, and E. Chuvieco, "Wildfires: Australia needs national monitoring agency," 2020.

[3] M. M. Boer, V. R. de Dios, and R. A. Bradstock, "Unprecedented burn area of australian mega forest fires," *Nature Climate Change*, vol. 10, no. 3, pp. 171–172, 2020.

[4] A. Singh, "Case study on 2019 australian bushfire," 04 2020.

[5] G. R. Van Der Werf, J. T. Randerson, L. Giglio, T. T. Van Leeuwen, Y. Chen, B. M. Rogers, M. Mu, M. J. Van Marle, D. C. Morton, G. J. Collatz *et al.*, "Global fire emissions estimates during 1997-2016," *Earth System Science Data*, vol. 9, no. 2, pp. 697–720, 2017.

[6] M. C. Stambaugh, L. D. Hammer, and R. Godfrey, "Performance of burn-severity metrics and classification in oak woodlands and grasslands," *Remote Sensing*, vol. 7, no. 8, pp. 10 501–10 522, 2015.

[7] E. Chuvieco, F. Mouillot, G. R. van der Werf, J. San Miguel, M. Tanase, N. Koutsias, M. García, M. Yebra, M. Padilla, I. Gitas *et al.*, "Historical background and current developments for mapping burned area from satellite Earth observation," *Remote Sensing of Environment*, 2019.

[8] C. H. Key and N. C. Benson, "Measuring and remote sensing of burn severity," in *Proceedings joint fire science conference and workshop*, vol. 2. University of Idaho and International Association of Wildland Fire Moscow, ID, 1999, p. 284.

[9] S. Trigg and S. Flasse, "An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah," *International Journal of Remote Sensing*, vol. 22, no. 13, pp. 2641–2647, 2001.

[10] M. Martín, I. Gómez, and E. Chuvieco, "Performance of a burned-area index (BAIM) for mapping Mediterranean burned scars from MODIS data," in *Proceedings of the 5th International Workshop on Remote Sensing and GIS Applications to forest fire management: fire effects assessment*. Universidad de Zaragoza, GOFC GOLD, EARSeL, 2005.

[11] L. Giglio, L. Boschetti, D. P. Roy, M. L. Humber, and C. O. Justice, "The Collection 6 MODIS burned area mapping algorithm and product," *Remote sensing of environment*, vol. 217, pp. 72–85, 2018.

[12] E. Chuvieco, J. Lizundia-Loiola, M. L. Pettinari, R. Ramo, M. Padilla, K. Tansey, F. Mouillot, P. Laurent, T. Storm, A. Heil *et al.*, "Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies," *Earth System Science Data*, vol. 10, no. 4, pp. 2015–2031, 2018.

[13] A. Bastarrika, E. Chuvieco, and M. P. Martin, "Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors," *Remote Sensing of Environment*, vol. 115, no. 4, pp. 1003–1012, 2011.

[14] D. Stroppiana, G. Bordogna, P. Carrara, M. Boschetti, L. Boschetti, and P. Brivio, "A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple Spectral Indices and a region growing algorithm," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 69, pp. 88–102, 2012.

[15] I. Alonso-Canas and E. Chuvieco, "Global burned area mapping from ENVISAT-MERIS and MODIS active fire data," *Remote Sensing of Environment*, vol. 163, pp. 140–152, 2015.

[16] A. A. Pereira, J. Pereira, R. Libonati, D. Oom, A. W. Setzer, F. Morelli, F. Machado-Silva, and L. M. T. De Carvalho, "Burned area mapping in the Brazilian Savanna using a one-class support vector machine trained by active fires," *Remote Sensing*, vol. 9, no. 11, p. 1161, 2017.

[17] F. Mouillot, M. G. Schultz, C. Yue, P. Cadule, K. Tansey, P. Ciais, and E. Chuvieco, "Ten years of global burned area products from spaceborne remote sensing—a review: Analysis of user needs and recommendations for future developments," *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, 2014.

[18] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*. Springer, 2018, pp. 270–279.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[21] D. Marcos, M. Volpi, B. Kellenberger, and D. Tuia, "Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 96–107, 2018.

[22] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 139–149, 2017.

[23] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 1–9.

[24] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.

[25] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, "A cloud detection algorithm for satellite imagery based on deep learning," *Remote sensing of environment*, vol. 229, pp. 247–259, 2019.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS International Journal of Geo-Information*, vol. 7, no. 4, p. 129, 2018.

[28] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal convolutional neural network for the classification of satellite image time series," *Remote Sensing*, vol. 11, no. 5, p. 523, 2019.

[29] A. Zulfiqar, M. M. Ghaffar, M. Shahzad, C. Weis, M. I. Malik, F. Shafait, and N. Wehn, "Ai-forestwatch: semantic segmentation based end-to-end framework for forest estimation and change detection using multi-spectral remote sensing imagery," *Journal of Applied Remote Sensing*, vol. 15, 2021.

[30] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais *et al.*, "Deep learning and process understanding for data-driven earth system science," *Nature*, vol. 566, pp. 195–204, 2019.

[31] M. M. Pinto, R. Libonati, R. M. Trigo, I. F. Trigo, and C. C. DaCamara, "A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 260–274, 2020.

[32] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A Deep Learning Approach for Burned Area Segmentation with Sentinel-2 Data," *Remote Sensing*, vol. 12, no. 15, p. 2422, 2020.

[33] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th International Conference On Machine Learning*, 2009, pp. 41–48.

[34] A. Ul-Hasan, F. Shafait, and M. Liwicki, "Curriculum learning for printed text line recognition of ligature-based scripts," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 1001–1005.

[35] "Fires in victoria destroy estimated 300 homes, former police chief to lead bushfire recovery victoria - abc news," https://www.abc.net.au/news/2020-01-06/bushfires-in-victoria-destroy-at-least-200-homes/11844292, Januray 2020.

[36] "Australia bushfires continue to wreak havoc – the state times," https://thestatetimes.com/2020/02/07/australia-bushfires-continue-to-wreak-havoc/, February 2020.

[37] "Australia fires: A visual guide to the bushfire crisis - bbc news," https://www.bbc.com/news/world-australia-50951043, January 2020.

[38] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[39] H. Fan, L. Zheng, C. Yan, and Y. Yang, "Unsupervised person re-identification: Clustering and fine-tuning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 4, 2018. [Online]. Available: https://doi.org/10.1145/3243316

[40] R. M. S. Bashir, M. Shahzad, and M. M. Fraz, "VR-PROUD: Vehicle Re-identification using PROgressive Unsupervised Deep architecture," *Pattern Recognition*, vol. 90, pp. 52–65, 2019.