Cloud and Aerosol Detection Using Spectral, Spatial and Temporal observations from GOES-R ABI

Yihan Fang[1]

A scholarly paper in partial fulfillment of the requirements for the degree of Master of Science

May 2023

Department of Atmospheric and Oceanic Science, University of Maryland

College Park, Maryland

Advisor: Dr. Chenxi Wang[2,3]

Dr. Xinzhong Liang[1]

[1]University of Maryland, College Park
[2]GESTAR-II, University of Maryland, Baltimore County
[3]NASA GSFC

Abstract

Four machine learning (ML) models for cloud and aerosol classification, utilizing spectral, spatial, and temporal features of the Advanced Baseline Imager (ABI) on board the geostationary satellite GOES-R, are developed and evaluated. The models include a Random Forest (RF) for extracting ABI spectral features, a Long Short-Term Memory (LSTM) for extracting spectral and temporal features, and a ConvLSTM-UNet model for extracting features from all dimensions. A new model that combines the RF and LSTM by leveraging the advantages of both models is also developed. Training data include 9 visible and infrared channels and 3 parameters calculated with specific channel combinations. Training datasets were manually labeled by domain experts, while test datasets were labeled using collocated CALIOP products. The accuracy scores in the training stage for RF, LSTM, and RF+LSTM exceed 0.99, while ConvLSTM-UNet achieves a slightly lower but still impressive accuracy of over 0.94. And the test accuracy scores of these models exceed 0.96, suggesting an excellent overall fit of these models. To further assess and compare their capabilities, some specific scenarios are employed, and their results are compared with RGB figures and ABI level-2 products. The findings unequivocally showcase the superior performance of ML models over traditional algorithms, emphasizing their effectiveness in accurately detecting and classifying cloud and aerosol components across diverse scenarios. The inclusion of temporal features in the LSTM and RF+LSTM models significantly improved their performance, particularly in distinguishing between thick dust, high-albedo surfaces, and clouds. Furthermore, one-year consecutive ABI data from five representative areas are analyzed to objectively assess the daily cloud fraction generated by the ML models, which closely aligns with both Terra and Aqua MODIS products. This agreement further confirms the robustness of the developed models. The comparison of spectral distributions between dust and cloud pixels, classified by RF and LSTM, demonstrates the limitations of the RF model, which solely relies on spectral features and can lead to misclassifications when confronted with objects exhibiting similar spectral distributions. However, the inclusion of temporal features significantly improves the RF model performance, enhancing its ability to discriminate between atmospheric components and the surface. These findings highlight the potential of ML models in satellite-based cloud and aerosol classification, contributing to advancements in remote sensing techniques. However, it is important to note the limitations observed in the ML models as well. RF encounters challenges in accurately identifying cloud and highly reflective objects, such as dust and snow. The ConvLSTM-UNet model has the potential to erroneously learn surface features, which can result in misclassifications. Furthermore, none of the ML models developed in this study demonstrate the ability to detect thin aerosols. To enhance the performance of the ML models in future research, we plan to collect data from diverse scenarios, incorporate land surface features into the models, and explore hybrid approaches that combine different techniques.

1. Introduction

With the rapid advancement of technology, remote sensing instruments have become indispensable tools for measuring the Earth's system. Satellites, in particular, enable observations of large-scale systems from the space. It has been demonstrated that clouds and aerosols play a crucial role in the Earth's radiation budget system while also remaining a significant source of uncertainty. Accurate detection and classification (DC) of clouds and aerosols using satellite observations are vital for meteorological studies, as even minor errors can have a significant impact on downstream retrieval products and scientific analysis (Wang et al., 2020). For example, access to cloud cover distribution substantially improves weather prediction and rainfall estimation systems (Torsum & Kwiatkowska, 1999). Furthermore, the Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol algorithm requires a cloud mask product as input, as it can only accurately retrieve aerosol optical depth (AOD) for "non-cloudy" pixels (Remer et al., 2005). Therefore, detecting and classifying clouds and aerosols represent vital pre-processing steps in many remote sensing algorithms, with significant implications for meteorological research and applications.

An increasing number of DC methods have been proposed for satellite instruments. For instance, Rossow and Schiffer (1991) proposed a cloud detection procedure for the International Satellite Cloud Climatology Project (ISCCP) based on a collection of visible and infrared radiance images. Vaughan et al. (2009) developed a selective, iterated boundary location (SIBYL) algorithm to retrieve clouds and aerosols in the earth's atmosphere for the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission. MODIS has also developed algorithms for cloud detection (Platnick et al., 2016; Frey et al., 2008; Ackerman et al., 2008) and aerosol types (Levy et al., 2013). Most DC methods primarily focus on individual pixels, which can easily result in the misclassification of pixels with similar spectral features. For example, operational MODIS cloud and aerosol products frequently misidentify optically thick dust layers as clouds. Consequently, large-impacts due to strong dust storms are missing (Marais et al., 2020). Another obvious weakness is that the traditional DC methods have difficulties in differentiating clouds from high reflective surfaces, such as sand in deserts, snow/ice, and human made infrastructures (Jeppesen et al., 2019). All these methods mentioned so far are single-time methods, which only need a single snapshot of observations. In contrast to these single temporal methods, Zhu and Woodcock (2014) developed a multitemporal mask (Tmask) algorithm which detects cloud, cloud shadow, and snow by using multitemporal images at the same location. Tmask needs 15 clear observations in each pixel to estimate a time series model, making it less applicable in places that are covered by snow and ice for a long period. As discussed earlier, traditional algorithms face significant challenges in accurately detecting clouds and aerosols using the current operational observation instruments, highlighting the need for more efficient methods that can improve accuracy.

Machine learning (ML) methods have shown great potential in improving the accuracy of cloud and aerosol detection. Recent applications using machine learning methods have already provided unprecedented improvements by utilizing spectral, spatial, and temporal information. Taravat et al., (2015) used multilayer perceptron (MLP) neural networks and support vector machine (SVM) to detect cloud in whole-sky images automatically. Ghasemian and Akhoondzadeh (2018) introduced two random forest (RF) based algorithms to detect cloud by incorporating spectral and textural features. Wohlfarth et al. (2018) explore convolutional neural network (CNN) and SVM

to distinguish 10 cloud types based on multispectral satellite imagery. Yu and Lary (2021) proposed an ensemble of different machine learning methods to detect cloud with channel information only. Liu et al. (2022) introduced both daytime and nighttime cloud detection algorithms on the basis of artificial neural network (ANN) and RF. In addition to detecting clouds, ML algorithms also have satisfying performance in detecting aerosols. Lee et al. (2021) developed 5 ML based algorithms to detect the occurrence of dust aerosols from daytime images under cloud-free conditions. Christopoulos et al. (2018) select chemical and physical features of particles with RF to differentiate aerosols. Marais et al. (2020) identify aerosols and distinct cloud regimes with CNN by leveraging the coherent spatial information in multispectral imagery.

The promising outcomes demonstrate the immense potential of machine learning (ML) methods for satellite-based aerosol and cloud detection and classification. Motivated by these successful applications, this paper presents the implementation of three classifiers: Random Forest (RF), Long Short-Term Memory (LSTM), and CNN U-net architecture. These algorithms are utilized to extract channel, temporal, and spatial features from satellite daytime observations separately, with the aim of improving the accuracy of detection and classification. Furthermore, we explore the benefits of combining different algorithms to further enhance the results. The paper is structured as follows: In Section 2, we provide a brief description of the individual ML algorithms and explain how these algorithms are combined in the experiment. Section 3 introduces the details of the data generated for training and validation purposes. In Section 4, we evaluate and discuss the performance of each model on the task of detection and classification (DC). Finally, in Section 5, we conclude the paper with a discussion on future work and potential research directions.

2. Methodology

This section presents a comprehensive overview of the individual ML models utilized in this study, namely Random Forest (RF), Long Short-Term Memory (LSTM), and U-net. Each model is described in detail, highlighting their unique characteristics and capabilities for extracting specific features from satellite observations. Additionally, we explain how these models are effectively combined to leverage their respective strengths and enhance the accuracy of detection and classification.

2.1 Random Forest (RF)

RF is an ensemble learning method that consists of a number of tree-structured classifiers to classification and regression (Breiman, 2001). Each individual tree in the RF votes for a prediction class and the prediction class with the maximum votes will be the final prediction class. Due to the large amount of randomly produced decision trees, one of the key advantages of RF is its robustness against overfitting and its ability to handle datasets of varying quality. Since RF generates a large number of decision trees using random sampling, it is less sensitive to the quality of the training data compared to other classification methods. This characteristic makes RF particularly suitable for classifying hyperspectral data, which often suffers from the curse of dimensionality and highly correlated features (Belgiu and Drăguţ, 2016). The remote sensing community has widely adopted RF due to its accuracy in classifying and detecting various phenomena. It has proven effective in handling complex and high-dimensional remote sensing datasets. RF has become a popular choice in remote sensing applications, including cloud and aerosol detection and classification. In this paper, we set the maximum depth of each tree to 20 and use a total of 120 estimators (i.e., decision trees) in the RF ensemble. These parameter values

are selected to optimize the performance of RF in classifying and detecting clouds and aerosols in satellite observations.

2.2 Long Short Term Memory (LSTM) networks

LSTM networks are an advanced version of Recurrent Neural Networks (RNN) that are capable of processing long-term dependencies, making it suitable for sequence prediction problems. They were designed by Hochreiter & Schmidhuber (1997) to solve the vanishing gradient and exploding problems of standard RNN. LSTM networks incorporate three gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information within the network. The input gate determines which elements of the input should be retained to update the memory, while the forget gate controls which information should be discarded. The output gate generates the output of the LSTM unit and determines the information to be passed on to the next LSTM unit. By learning which information to keep or discard, LSTM networks can effectively propagate relevant information through sequences and make accurate predictions.

With these advanced properties, LSTM has been widely used in meteorological studies, such as predicting sea surface temperature (Zhang et al., 2017; Yang et al., 2018), estimating the concentration of air quality pollution indicators (Kim et al., 2019; Wu and Lin, 2019; Seng et al., 2021; Sun et al. 2022), and change detection, which analyzes and quantifies temporal changes at the same geographical location (Wang et al., 2015), in remote sensing images (Jing et al., 2020). Change detection can also be applied for DC since the temporal variations of atmospheric components are different. Thus, LSTM shows a substantial potential to differentiate cloud and aerosol by detecting the pixel-based spectral changes of satellite data. In this paper, LSTM is employed to extract spectral and temporal features for distinguishing clouds and aerosols. The structure of the LSTM network used in this study is depicted in Figure 1, providing a visual representation of the network architecture.

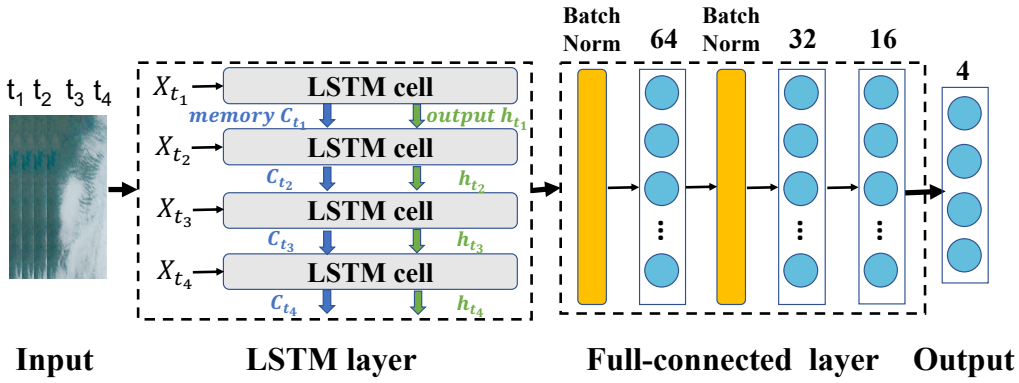

Figure 1. Structure of LSTM model

2.3 RF+LSTM

In this study, the limitations of both RF and LSTM in differentiating cloud and aerosol are addressed by combining the two algorithms. RF excels in scenarios where there are distinct differences in channel information but struggles when faced with objects that exhibit similar spectral features, such as clouds and thick aerosols. On the other hand, LSTM is adept at distinguishing similar objects by capturing their temporal variations, but its effectiveness diminishes when dealing with rapidly changing phenomena like clouds.

To overcome these limitations and enhance the performance of cloud and aerosol differentiation, a novel approach that combines RF and LSTM is proposed. This approach takes advantage of the unique strengths of each algorithm. Specifically, incorporate the LSTM-predicted probabilities of slowly varying objects, such as dust and clear air, as additional parameters in RF. By integrating these probabilities into RF, its ability to discriminate between clouds and thick aerosols is expected to be improved. By combining RF and LSTM, the complementary capabilities of both algorithms are leveraged, resulting in enhanced accuracy and robustness in cloud and aerosol detection from satellite observations. This integrated approach allows us to exploit the benefits of temporal variation analysis provided by LSTM while utilizing the superior feature extraction capabilities of RF. Through this combination, we aim to address the limitations of each algorithm and achieve improved performance in differentiating clouds and aerosols.

2.4 ConvLSTM-UNet

The U-Net architecture, initially introduced for biomedical image segmentation, has gained significant recognition in semantic segmentation tasks and pixel-based classification. Ronneberger et al. (2015) proposed U-Net as a convolutional network architecture that aims to classify each pixel in an image into a specific class, unlike traditional CNNs that focus on the final classification result. The U-Net architecture comprises two paths: the contracting path (encoder) and the expansive path (decoder). The encoder network functions as a typical convolutional neural network, extracting features and learning abstract representations from the inputs. The decoder network projects the features learned by the encoder onto the pixel space, reconstructing a dense classification output from lower to higher resolution. The skip connections play a crucial role in preserving information by combining data from both encoder and decoder blocks, preventing feature loss. The bridge connects the contracting and expansive paths, facilitating the flow of information. U-Net has demonstrated superior performance compared to its predecessors, achieving accurate segmentation with minimal training images, reduced information loss, and lower computational costs. Given these advantages, U-Net is selected as the backbone architecture in this study to differentiate pixel-based clouds and aerosols.

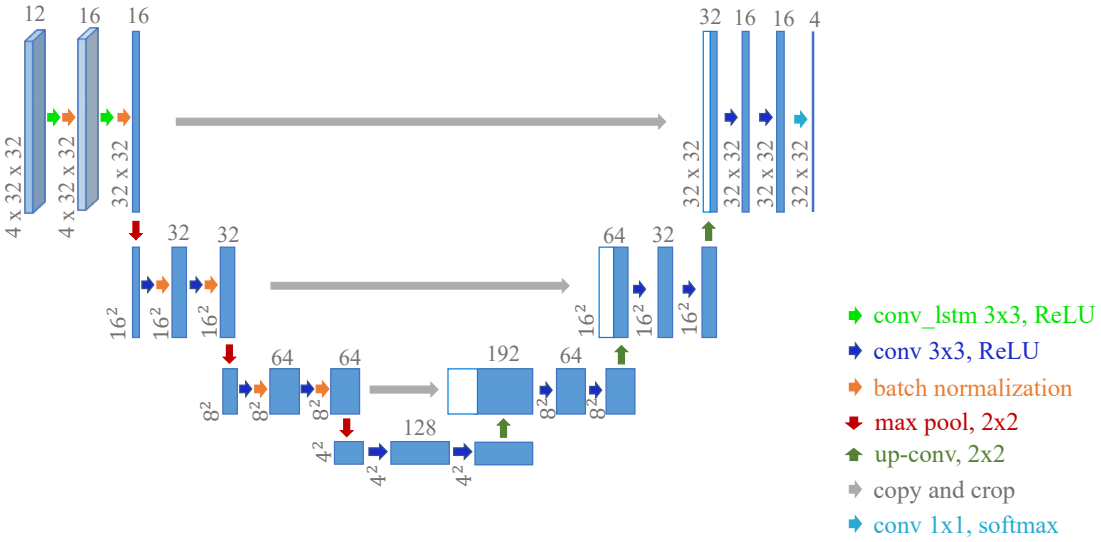

Figure 2. Structure of ConvLSTM-UNet model

As is mentioned above, the LSTM has proven power for extracting temporal features, but it's powerless for spatial data. Indeed, to address the limitations of LSTM in handling spatial data, Shi

et al. (2015) proposed an extension called ConvLSTM. ConvLSTM incorporates convolutional structures in both the input-to-state and state-to-state transitions, effectively replacing internal matrix multiplications with convolution operations. This modification enables ConvLSTM to process sequential images, where the input data dimension is 3D, unlike LSTM that operates on 1D vectors. In the context of differentiating between cloud and aerosol pixels, the ConvLSTM-UNet network architecture is employed to leverage the strengths of both ConvLSTM and U-Net. The ConvLSTM layers within the model enable the processing of sequential images, allowing for the extraction of temporal information. On the other hand, the U-Net structure facilitates pixel-based classification and preserves spatial information.

The ConvLSTM-UNet model consists of two main paths: the encoding path and the decoding path. The encoding path comprises several convolutional layers followed by pooling layers, which progressively extract hierarchical features from the input image. In the ConvLSTM-UNet architecture, the first encoding block is replaced with ConvLSTM layers to enable sequential image processing. The decoding path consists of several upsampling layers followed by convolutional layers, which gradually reconstruct the high-resolution classification map. Skip connections are included in the architecture to preserve spatial information and prevent feature loss. These skip connections combine information from corresponding encoder and decoder blocks, allowing for the fusion of multi-scale features. Ultimately, the output of the ConvLSTM-UNet model is a pixel-based classification map that assigns each pixel in the input image as either cloud or aerosol. The network architecture and its components are illustrated in Figure 2, providing a visual representation of the ConvLSTM-UNet structure.

3. Data

3.1 Training and validation datasets

Generating accurately labeled training data is crucial for training machine learning models. In this study, a geostationary instrument, the advanced Baseline Imager (ABI), onboard NOAA's GOES-R is used as it provides continuous observations at a 10-minute interval for the full disk. It captures imagery of the Earth with 16 different spectral bands, including visible, near infrared, and infrared channels. The ABI offers improved capabilities compared to previous GOES imagers, including more spectral channels, faster image repetition rates, and higher spatial resolutions (Schmit et al., 2017). These features make the ABI well-suited for detecting various elements on Earth's surface, such as clouds, water, moisture, and smoke.

For the purpose of this study, the ABI data are manually categorized into four distinct classes: clear, cloud, dust, and smoke. Clear means no aerosol and cloud scenario, cloud includes both ice and liquid cloud types, and aerosol types are divided into dust and smoke categories. Labeling training data for the RF, LSTM, and RF+LSTM models is relatively straightforward as these methods do not require labels for surrounding pixels. We manually generated 80,000 pixel-level labels for each of the models. However, the training dataset for the ConvLSTM-UNet model differs from the other three models. It is constructed using 1000 manually selected patches of size 32×32, with approximately 800 patches used for training and the remaining 200 for validating the accuracy of the trained model. Labeling pixels for the ConvLSTM-UNet model presents a challenge as it requires grouping pixels together (i.e., 32×32 pixels) for labeling. To overcome this challenge, labels from the other three models (RF, LSTM, and RF+LSTM) are manually selected as reference points. These models have demonstrated superior performance in pixel

classification, making their predicted labels suitable as guides. To ensure the accuracy of the selected labels, correlated RGB images are plotted as additional references. The labels are chosen from the classification results of the three models only when they align closely with the RGB image, with a matching rate of over 90%. Furthermore, to ensure the accuracy of the ConvLSTM-UNet model's labels, the selected labels are also verified using NASA's Worldview web platform. This meticulous manual selection process serves to verify and validate the predicted labels, ensuring their accuracy and bolstering the reliability of the training data for the ConvLSTM-UNet model.

3.2 Test datasets

Accurately labeled test datasets are indeed crucial for evaluating the performance of machine learning models. In this study, to ensure the reliability of the evaluation, test datasets were generated using cloud and aerosol labels provided by active remote sensing instruments, specifically the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard the CALIPSO satellite. CALIOP is a polarization-sensitive lidar that operates at two wavelengths (532nm and 1064nm). It is capable of providing high-resolution vertical profiles of global clouds and aerosols in the lower troposphere and lower stratosphere. CALIOP employs sophisticated algorithms to detect and retrieve the microphysics and optical properties of these atmospheric components, representing a significant advancement over previous lidar instruments (Winker et al., 2009). The cloud and aerosol labels obtained from CALIOP are considered to be more reliable compared to labels from other instruments. These labels serve as a reference for evaluating the performance of the machine learning models in this study. To construct the test datasets, ABI data was utilized along with the corresponding labels provided by CALIOP for collocated observations. The ABI data contains multi-channel spectral information that enables the differentiation of various elements on Earth's surface, including clouds and aerosols.

Table 1 Summary of Training and Testing datasets

Models	Training/Validation Datasets	Testing Datasets
RF LSTM	Manually picked clear/cloud/dust/smoke cases and labels (samples: 80,000 pixels)	ABI 32x32 patches whose center pixels are labeled by CALIPSO. (Samples: 15312 patches)
RF + LSTM	Manually Picked Cases and Labels, LSTM clear and dust category probabilities (samples: 80,000 pixels)	
ConvLSTM-UNet	Manually Picked patches with labels from RF/RF+LSTM/LSTM with higher probability (samples: 1,000 patches)	--

Table 1 summarizes the training and testing datasets used for the four models in this study, indicating the sources of the data and the corresponding labels. By using the CALIOP observations in conjunction with the ABI data, the study ensures the availability of accurately labeled test datasets, which facilitates a reliable evaluation of the performance of the machine learning models. Note that for the ConvLSTM-UNet model, the test dataset cannot be generated from CALIOP

labels. Therefore, the evaluation of its classification results is conducted by comparing them with RGB images.

3.3 Quality assurance of labeled data

The quality of training and testing datasets is indeed crucial for training machine learning models. In this paper, the manually labeled data were collected by interacting with the RGB plotting of ABI data. The red, green, and blue channels of the ABI data were represented by the 1600nm, 860nm, and 470nm channels, respectively. To ensure the accuracy of the labels, the researchers utilized NASA's Worldview (https://worldview.earthdata.nasa.gov/) interface as a reference. The Worldview web interface provides a comprehensive record of the changes in objects over time, making it easier to identify ambiguous pixels, such as thin clouds and thick aerosols or the distinction between snow and clouds in polar regions. For LSTM-related models, time series datasets were constructed. However, the labeling of these time-series pixels presented a challenge. To address this issue and avoid inconsistencies in the data, it was essential to ensure that each pixel belonged to a specific category throughout the entire time series. To achieve this, images with a size of 32 × 32 pixels, centered around the selected pixels, were plotted and used to select qualified cases. Only these qualified cases were used to construct the training datasets, ensuring the quality of the training database and avoiding confusion in the model.

By carefully selecting and labeling the data using the ABI RGB plotting and referencing the Worldview web interface, as well as addressing the challenges of labeling time-series pixels, this study aimed to enhance the quality and accuracy of the training and testing datasets, thereby improving the overall performance of the machine learning models.

3.4 Input and Output of the ML Models

In this study, a total of 12 parameters comprising 9 channel observations and 3 parameters have been carefully selected as input spectral features. This selection includes the incorporation of both shortwave reflection and longwave emission, which aims to capture a wide range of features relevant to the target categories. To ensure consistency in the analysis, the resolutions of the channels used for collecting data (which vary between 0.5km, 1km, and 2km) are standardized to a uniform resolution of 1km. The three indexes employed are BTD, NDVI, and NDAI:

$$BTD = 8.4\text{um} - 11.2\text{um} \tag{1}$$

$$NDVI = (0.86\text{um} - 0.67\text{um})/(0.86\text{um} + 0.67\text{um}) \tag{2}$$

$$NDAI = -10log_{10}(0.46\text{um} - 2.2\text{um}) \tag{3}$$

which account for surface types and attempt to mitigate their influence on the classification process. As previously discussed, the study examines four distinct models, each of which necessitates specific input requirements. In particular, the RF only considers 12 spectral features as is discussed above. The LSTM and RF+LSTM models incorporate temporal variations by incorporating the time dimension as an additional factor. The input data for these models encompasses consecutive 4 10-minute interval observations. On the other hand, the ConvLSTM-UNet model takes into account spatial characteristics in addition to temporal variations. Instead of considering individual pixels, the input for the ConvLSTM-UNet model comprises 32 × 32 pixel patches, with a resolution of 1 km.

The outputs of RF, LSTM, and RF+LSTM models consist of four probabilities denoting the likelihood of being categorized as clear, cloud, dust, or smoke. To determine the final label, only the category with the highest probability is selected. In contrast, the output of the ConvLSTM-UNet model differs in nature. It produces 32 × 32 pixel patches, each containing four possibilities. Similar to the aforementioned models, the final label is determined by selecting the option with the highest probability. However, only the 16 × 16 pixels situated at the center of the patches are considered to be reliable. This approach is adopted because the encoding and decoding processes involved in the U-Net structure may lead to the loss of edge information. The detailed specifications of the inputs and outputs for all models can be found in Table 2.

Table 2 Models Input and Output

<table>
<tr><th>Input Dimensions</th><th>Spectral Dimension</th><th>Temporal Dimension</th><th>Spatial Dimension</th><th>Output Dimensions</th></tr>
<tr><td>RF (1D)</td><td rowspan="4">Reflectance: 0.46, 0.67, 0.86, 1.37, 1.6, and 2.2 um
BT: 8.4, 11.2, and 13.3 um
BTD
NDVI
NDAI</td><td>N/A</td><td>N/A</td><td rowspan="3">Single Pixel Category (Clear/Cloudy/Dust/Smoke)</td></tr>
<tr><td>LSTM (2D)</td><td rowspan="3">4 consecutive observations, 10-mins interval</td><td rowspan="2">N/A</td></tr>
<tr><td>RF + LSTM (2D)</td></tr>
<tr><td>ConvLSTM-UNet (4D)</td><td>32 × 32 (1km) patches</td><td>Category for a group of pixels 16 × 16</td></tr>
</table>

4. Results and discussion

During the training stage of the four models, a random selection process is employed to allocate 80% of the data for training, while the remaining 20% of are reserved for validation. To ensure the reliability and robustness of the models, a 5-fold cross-validation approach is employed for sample selection during training and validation in this study. This repetition allows for the validation of the models' performance across multiple iterations, providing a comprehensive evaluation of their robustness and generalizability. The training-validation samples are randomly chosen to maintain similar pixel type distributions between the training and validation sets.

4.1 Training and test accuracy

The evaluation of classification models is essential to assess their reliability and applicability in various scenarios. In this study, an accuracy score and a confusion matrix are employed as evaluation metrics for the ML models.

The accuracy score is a commonly used metric in classification models, which measures the proportion of correct predictions (both true positives and true negatives) out of the total predictions made by the model. It is calculated using the formula:

$$Accuracy\ Score = (TP + TN)/(TP + TN + FP + FN), \quad (4)$$

where TP (True Positive) and TN (True Negative) represent correctly predicted positives and negatives, respectively, while FP (False Positive) and FN (False Negative) represent incorrectly predicted positives and negatives, respectively. Table 3 displays the training and test accuracy scores for all the models developed in this study. The notably high accuracy scores achieved for

both validation and testing phases indicate the excellent performance of these models. Figure 3 showcases the corresponding validation confusion matrices, which further illustrate the satisfactory accuracies observed for each class, aligning with our expectations.

Table 3 Validation and test accuracy scores

Model	RF	LSTM	RF+LSTM	ConvLSTM-UNet
Validation accuracy score	0.99	0.99	0.99	0.94
Test accuracy score	0.96	0.96	0.96	--

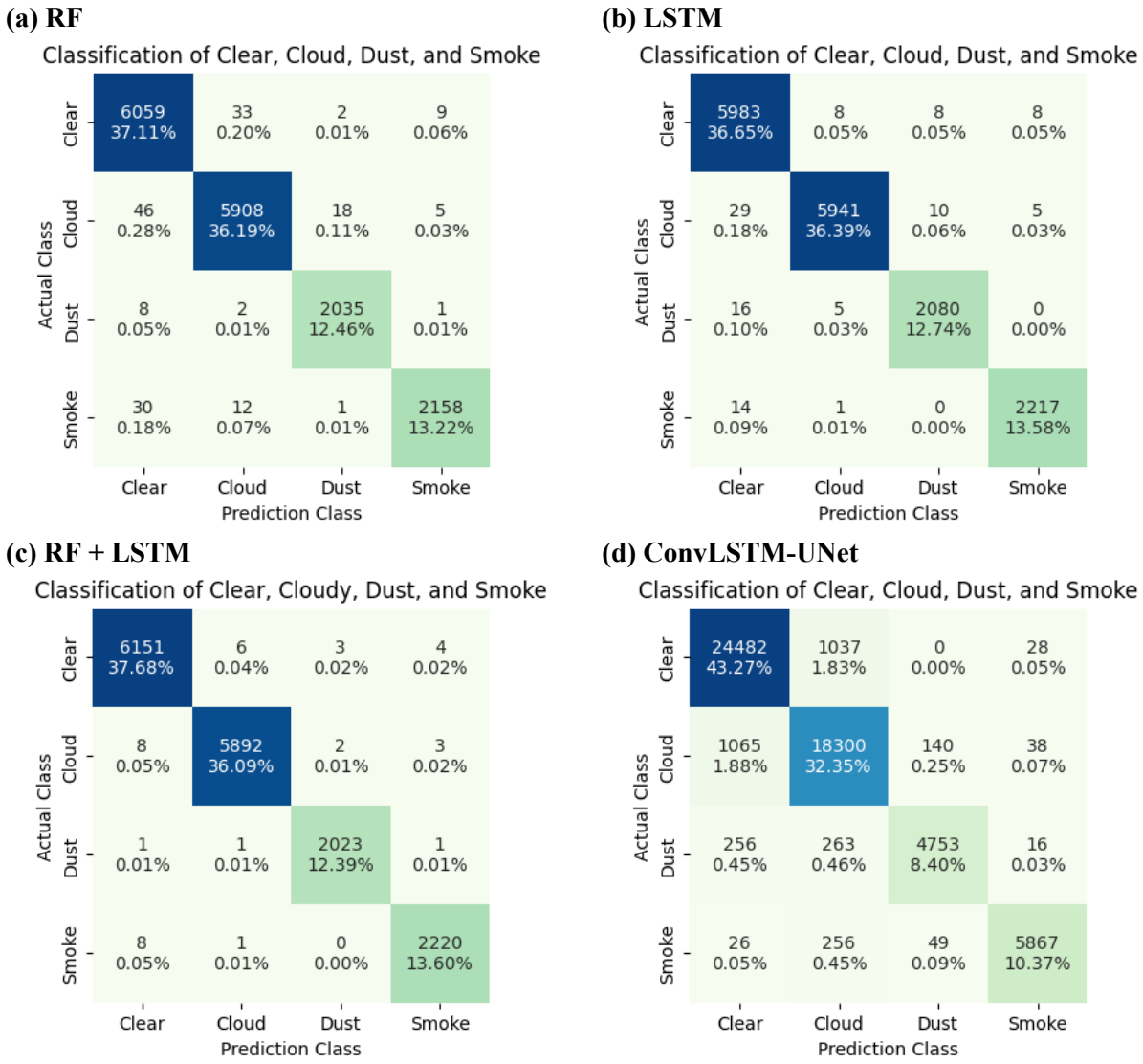

Figure 3. The confusion matrixes of four ML models in four scenarios. The diagonal elements indicate the correctly classified cases.

When a LSTM layer is incorporated, such as in the LSTM and RF+LSTM models, fewer clear pixels are wrongly predicted as clouds compared to the RF model. The misclassification rate is 0.05% and 0.04% for the LSTM and RF+LSTM models, respectively, whereas it is 0.20% for the

RF model. Additionally, the inclusion of LSTM layers reduces the error rate of classifying cloud pixels as clear. These results indicates that the LSTM layer is effective at distinguishing objects with different rates of temporal variation. Another notable improvement is the reduced misclassification of smoke pixels as clear pixels. This suggests that even when there is minimal difference in channel features between clear and smoke pixels, they can still be differentiated through temporal changes and spatial features extracted by the LSTM layers and U-Net structure, respectively. The training and validation accuracy scores demonstrate the excellent performance of the four models. However, these results alone cannot validate the models' robustness in real-world scenarios.

4.2 Case studies

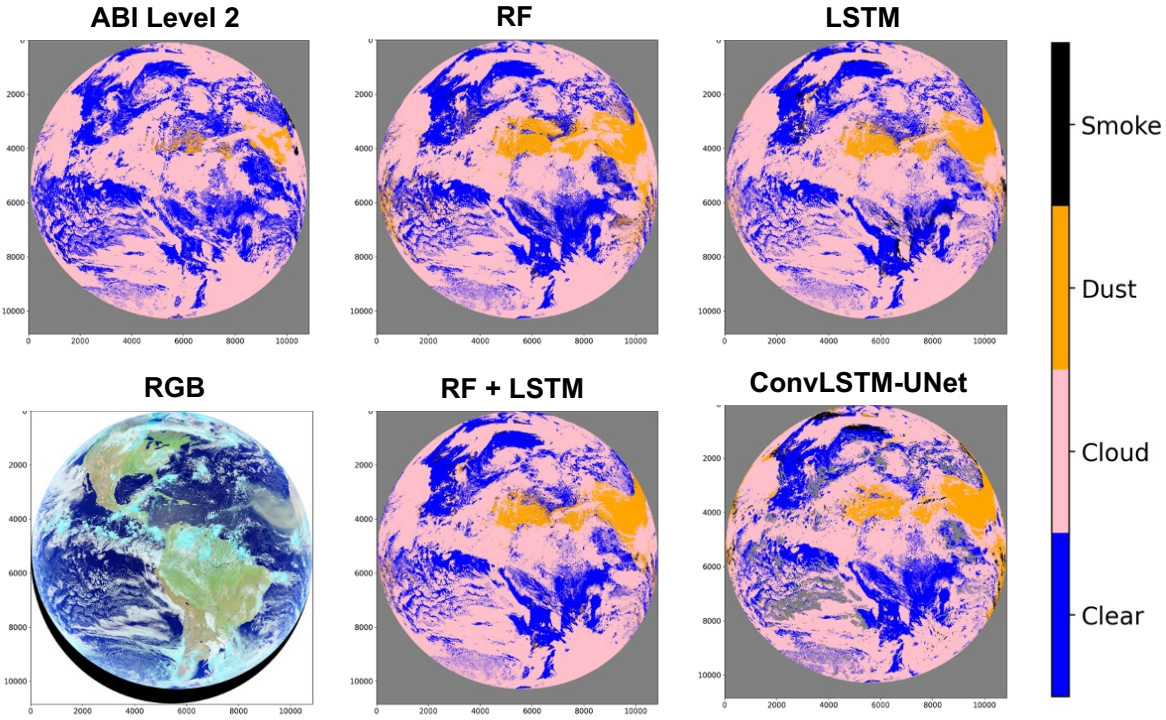

Figure 4. GOES-R 2020-06-28 16:40 (UTC) full disk classification results generated from four ML models. ABI Level-2 products and RGB image are also shown for comparison.

To assess the practical performance of the four models, a range of scenarios is employed for testing purposes. These scenarios are carefully chosen from NASA's Worldview Web platform, ensuring a diverse and representative selection of cases. The RGB images used for testing are constructed using three different channels: 1.6um, 0.86um, and 0.47um, which are assigned as the red, green, and blue channels, respectively. In addition to the RGB images, NOAA ABI level-2 products, combination of Aerosol Detection product (The aerosol mask, which indicate the presence of either smoke or dust) and Clear Sky Mask product (a binary cloud mask that identifies covered pixels as clear or cloudy), are plotted for comparison purposes.

Figure 4 shows the classification results from different ML models and NOAA ABI products of a full-disk image on June 28, 2020 (16:40 UTC). The results from both the NOAA ABI product using traditional algorithms and our ML models demonstrate commendable performance when comparing with the RGB image. However, some discrepancies can be observed in certain areas.

One notable difference is observed in western Africa, an area heavily affected by thick dust. In this region, both the RF model and the ABI product incorrectly identify the dense aerosol as clouds. However, when temporal features are incorporated through the use of LSTM, RF+LSTM, and ConvLSTM-UNet models, the classification of dust over the Atlantic area shows improvement. This promising outcome highlights the value of incorporating temporal features learned by these models. Further investigation and validation of specific cases are currently underway to confirm the advancements offered by these machine learning models. In addition to the full disc image, we compared a coupled of featured cases, namely, Uyuni Salt Flat, mixed cloud and thick dust plume, thin and thick smoke, are discussed below.

4.2.1 Uyuni Salt Flat

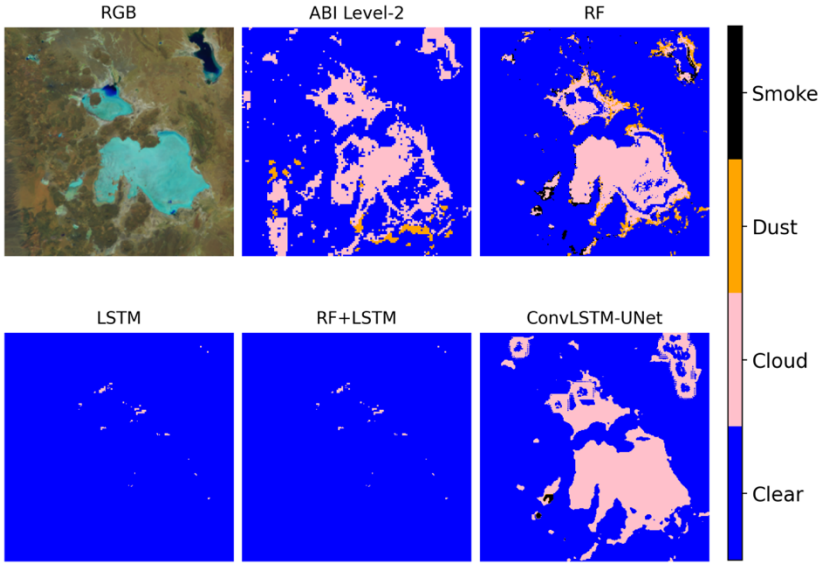

Figure 5. Similar to Fig. 4, but a Salt Flat case captured on 2020-06-20 at 17:30 (UTC).

Figure 5 showcases an image of the Uyuni Salt Flat in southwest Bolivia on 2020-06-20 at 17:30 (UTC). The ABI level-2 product, the RF and ConvLSTM-Unet models classified the salt flat and the edges of a nearby lake (located northeast of the images) as a cloud due to its similar bright color and shape. However, the incorporation of temporal features in the LSTM and RF+LSTM models resolved the classification issue as anticipated. These results not only highlight the advanced performance of the LSTM and RF+LSTM models but also emphasize the significance of incorporating temporal features for accurately identifying high-reflective objects and clouds. Despite the outstanding performance, both the LSTM and RF+LSTM models still exhibited misclassification of some pixels at the edge of the salt flat. It appears that edge detection remains a significant challenge for all the algorithms examined in this study.

4.2.2 Mixed cloud and thick dust plume

The second case is a mixed strong dust and cloud scenario close to the west shore of Africa on the eastern Atlantic Ocean, 2020-06-18 at 16:40 (UTC). Figure 6 reveals that all the ML models successfully identified clouds with cold toppings in the strong dust event, whereas the ABI Level-2 products erroneously classified most dust pixels as clouds. Additionally, models without a convolutional layer (RF, LSTM, RF+LSTM) provided more detailed information about the detected objects. This finding suggests that the inclusion of convolutional layers may result in some information loss. Results of this scenario underscore the superiority of all present ML models over the traditional algorithm in accurately detecting and classifying objects.

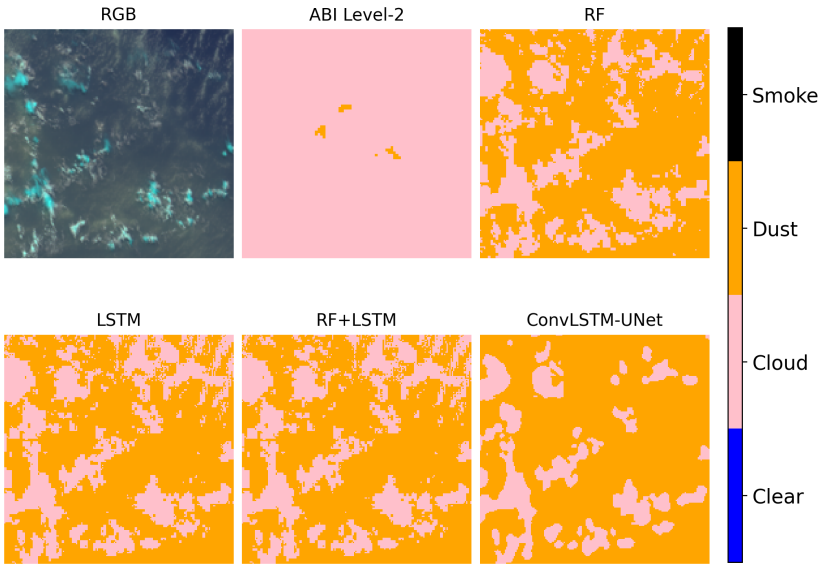

Figure 6. Similar to Fig. 4, but a mixed cloud and thick dust plume case captured on 2020-06-18 at 16:40 (UTC).

4.2.3 Thick smoke

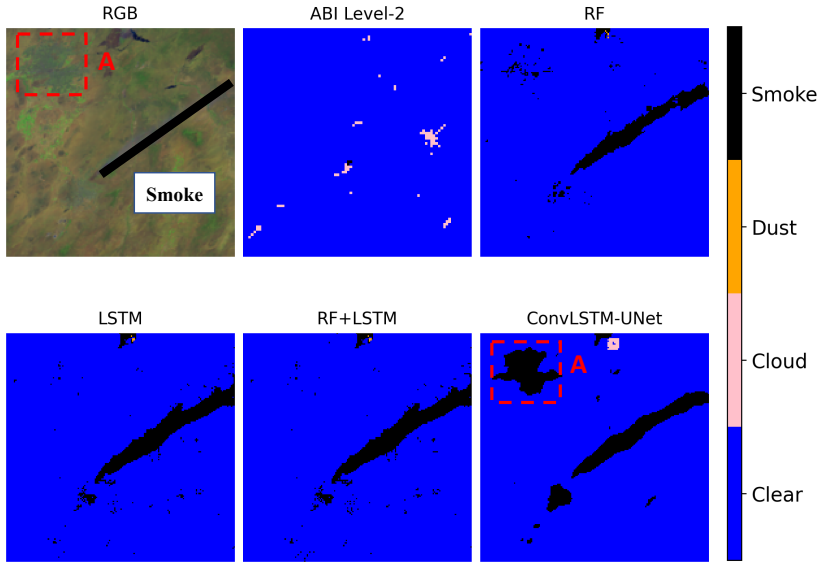

Figure 7. Similar to Fig. 4, but a thick smoke case captured on 2020-06-18 at 16:40 (UTC)

Figure 7 presents an image of a thick smoke event captured on 2020-06-18 at 16:40 (UTC), in western USA. The RGB channels combination used in this study does not provide a clear depiction of the smoke case due to the interference of surface features. The path of the smoke is marked with a thick black line in the image for reference. As anticipated, all the ML models successfully detected the path of heavy smoke, while the ABI Level-2 products inaccurately classified the smoke as clear and cloud. However, Box A in Figure 7 reveals an erroneous result produced by the ConvLSTM-UNet model. Upon comparison with the RGB image, it becomes evident that the ConvLSTM-UNet model mistakenly identified the surface texture as smoke. This observation suggests that while the U-Net structure aids in learning spatial features, it can be prone to confusion

when confronted with complex surface textures. To enhance model performance, future improvements could involve incorporating surface features.

All the results showed so far indicate the advancement of machine learning models, but it is found that they do experience some challenges in some scenarios.

4.2.4 Thin smoke

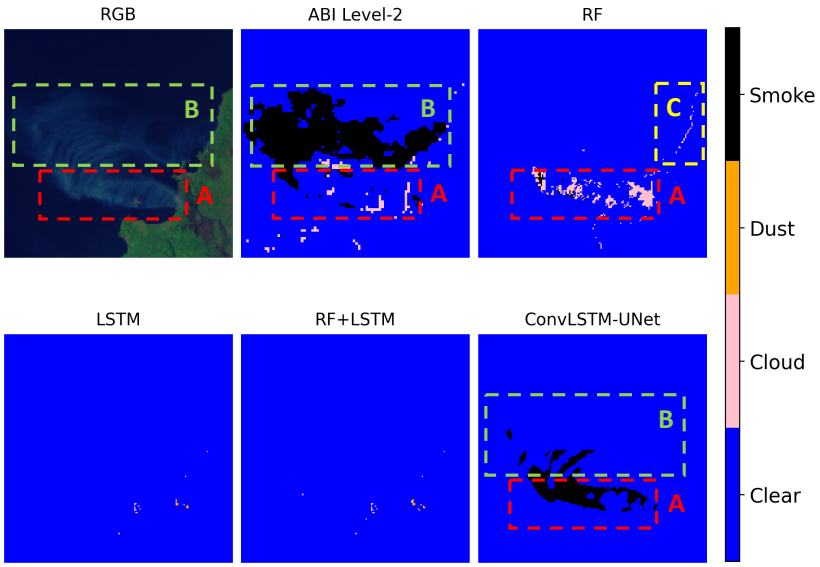

Figure 8. Similar to Fig. 4, but a thin smoke case captured on 2023-02-17 at 21:50 (UTC)

Figure 8 displays an image of a thin smoke event captured on 2023-02-17 at 21:50 (UTC), near the west coast of South America. It is evident that none of the products accurately detect the entire smoke areas. Box A in represents the main branch of the smoke event, which appears relatively thicker. Only the ConvLSTM-UNet model correctly identifies the shape of the smoke in this region, although some pixels are still misclassified as clear. The RF model predominantly classifies most pixels in this area as cloud, which aligns with the results obtained from the ABI Level-2 products. Moving on to Box B, the spread of the smoke is thinner compared to Box A. The ABI Level-2 products exhibit excellent performance in detecting this thin smoke, whereas all the ML models fail to capture this event, including the ConvLSTM-UNet model, which performes well in Box A. Particularly disappointing is the fact that the LSTM and RF+LSTM models, which emphasize temporal features, completely miss this thin smoke event. This outcome suggests that traditional algorithms excel at identifying thin smoke, and spatial information plays a critical role in machine learning models' ability to detect smoke. Lastly, Box C highlights the Random Forest model misclassifying the pixels at the edge of the continent as clouds, a recurring edge-related error observed in the salt flat case (Fig. 5) as well.

4.3 Statistical analysis

In order to conduct a thorough and unbiased evaluation and comparison of DC models, we collected continuous observations at three-day intervals throughout the year 2020 from the ABI and classified them using machine learning models built in this study. To provide a reference point, the daily mean cloud fraction products from MODIS on both Terra and Aqua are used. Rather than generating a full-disk prediction, which would be resource-intensive and time-consuming, only five typical areas (as shown in Figure 9) are selected and classified every hour during daytime (i.e., when the solar zenith angle is less than 80 degrees).

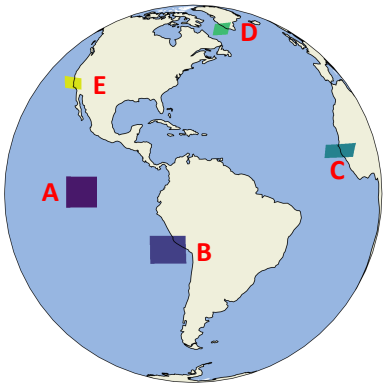

Figure 9. Five selected representative areas used for statistical analysis. These areas encompass a diverse range of scenarios discussed in the study.

Area A is situated at the Inter-Tropical Convergence Zone (ITCZ) region, which experiences an abundance of cirrus and convective clouds throughout the year. Area B is located along the mid-west coast of South America, while area E is situated on the west coast of California. Both areas experience numerous smoke events over a large area of semi-persistent marine boundary layer clouds, especially during summer and autumn. Area C, located near the Sahara Desert, is characterized by frequent dust events that occur throughout the year, with a higher occurrence during the late spring to mid-fall period. Area D is situated near Greenland, where ice cover exists throughout the year and the daytime is relatively short during the winter. Given that the detection targets are present in all five areas, analyzing the statistical prediction results for these regions would be highly meaningful. And some interesting results are shown in this section.

4.3.1 Cloud fraction comparison with MODIS product

Figure 10 illustrates a comparison of cloud fraction variations derived from MODIS L3 Daily products and the predictions of machine learning models for five specific areas. The Terra traverses each of these areas at 10:30 AM (local time), while Aqua passes at 1:30 PM (local time), approximately. For comparison purposes, ML model classification results for each area at the corresponding local time are utilized. The MODIS cloud fraction for each area is computed using the "Cloud Fraction Day Mean" and "Cloud Fraction Day Pixel Counts" extracted from MOD08_D3 (MYD08_D3 for Aqua), represented by the gray line.

As depicted in Figure 10, all the ML models developed in this study exhibit satisfactory overall performance, displaying close correspondence with both Terra and Aqua MODIS products. However, noticeable discrepancies are observed in the comparison results for Area C. Area C is situated in close proximity to the Sahara Desert, an area characterized by a persistent presence of clouds and dense aerosols, in particular from Summer to mid-Autumn. It is widely recognized that MODIS products may not consistently discern optically thick aerosols and cloud types (Marais et al., 2020). Consequently, it is highly probable that the MODIS product misclassifies the thick dust as clouds in Area C, thereby justifying the higher cloud fraction values derived from the MODIS product. To further quantify the comparison between MODIS and ML model products in each area, the correlation coefficient between the MODIS product and the result obtained from each machine learning model is calculated and presented in Figure 11.

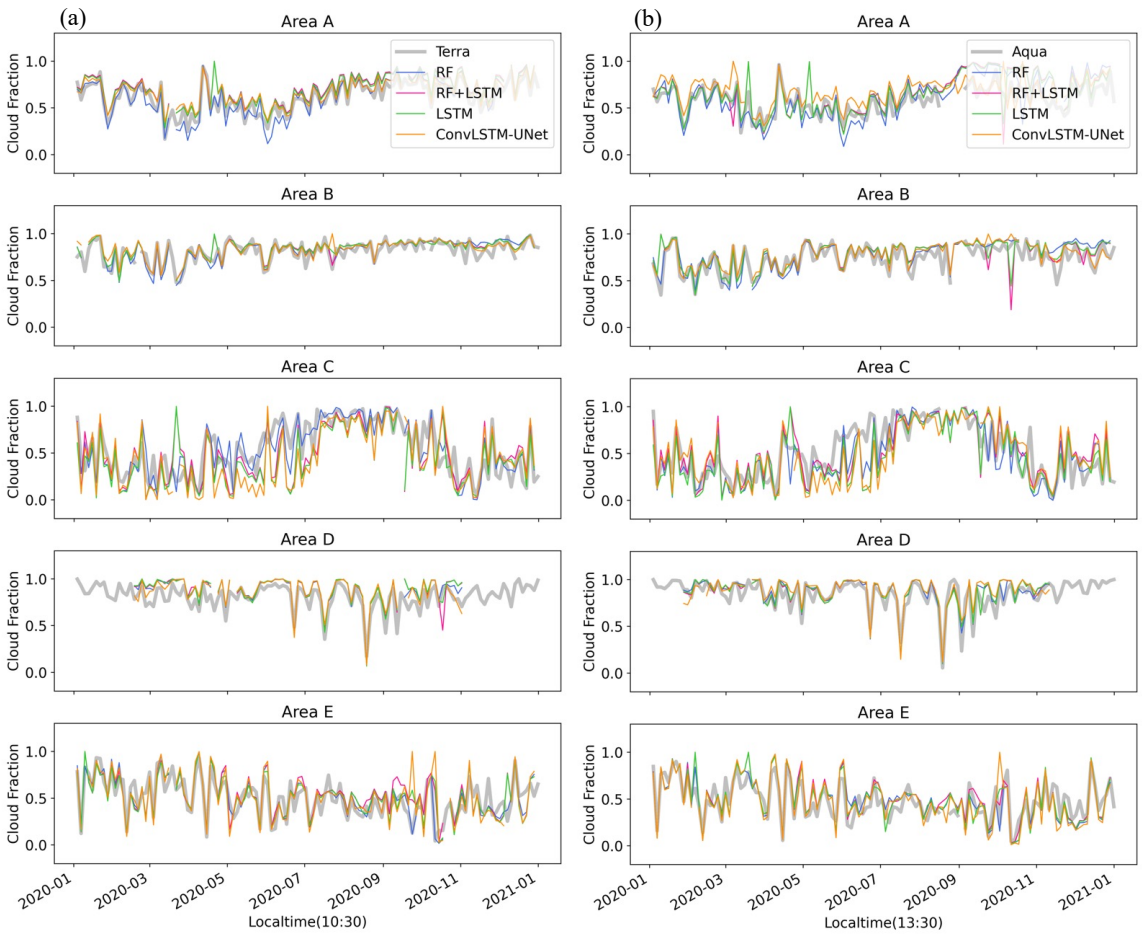

Figure 10. Comparison of cloud fraction products, which are generated from ML models classification results and MODIS Level-3 products (a. Terra and b. Aqua), separately.

It is evident that all ML models and MODIS products (both Terra and Aqua) exhibit the relatively poor agreement in Area C, except RF. It is expected that the LSTM or RF+LSTM models would demonstrate superior performance, as evidenced by the case studies conducted. However, it is observed that the RF model yields the closest match with the MODIS products. There are two reasons that can be attributed to this outcome. Firstly, while LSTM layers indeed enhance the detection of dust and clear pixels, their performance is adversely affected when it comes to cloud identification. The LSTM layer enables the model to learn features from all inputted timesteps in order to predict the label of the last timestep. Consequently, if a pixel was contaminated by a cloud in previous timesteps, it will be erroneously labeled as a cloud. This issue has the most significant impact on the classification of clear pixels. Hence, the adverse impact of LSTM layers on cloud detection contributes to the discrepancy observed. The second reason lies in the fact that both the traditional algorithm employed by MODIS and the RF algorithms solely utilize spectral information to generate labels for each pixel. It has already been demonstrated that the inclusion of additional temporal features alters the classification results, thereby resulting in discrepancies between the other three models and RF, and consequently, disparities with the MODIS products.

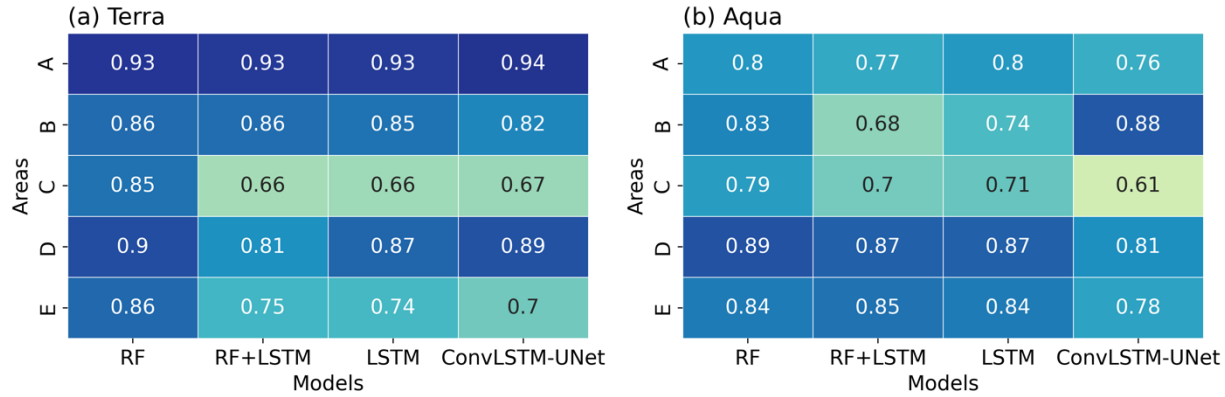

Figure 11. The correlation coefficients of the daily cloud fraction between the MODIS Level-3 product from Terra and Aqua satellites and the corresponding results obtained from each ML model.

4.3.2 Spectral distribution comparison between cloud and dust

As discussed earlier, RF is more prone to identify thick dust as cloud, while the LSTM and LSTM+RF algorithms are more accurate in thick dust detection. Consequently, an investigation is conducted into the channel distribution of dust and cloud pixels classified by RF and LSTM. These pixels were collected from one year of consecutive observations in area C, which is located near the Sahara Desert and experiences a high presence of clouds and thick dust. Figure 12 illustrates the statistical results obtained from RF and LSTM, with a selection of 80,000 pixels for each distribution. Specifically, we focused on the pixels identified as dust by LSTM but mistakenly classified as clouds by RF.

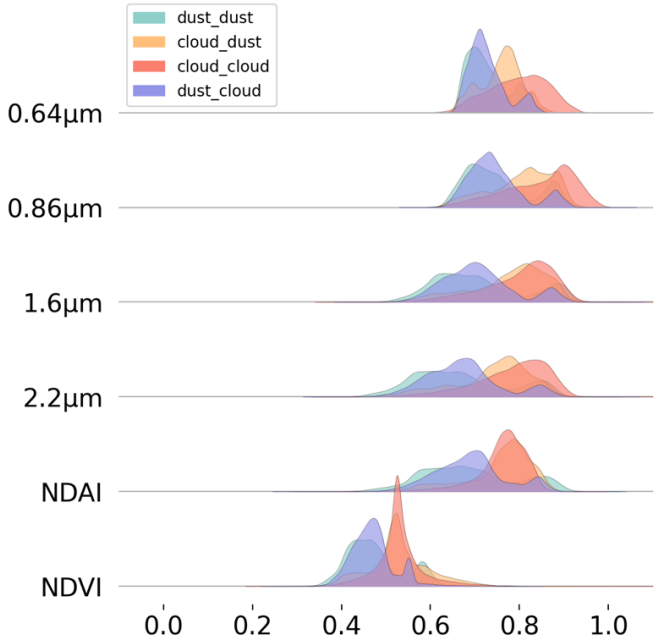

Fig 12. Comparison of cloud and dust spectral features classified by RF and LSTM models. The data is collected from the west coast of Africa, which is characterized by a significant presence of clouds and thick dust.

The spectral distribution plot in Figure 12 provides valuable insights into the performance of the Random Forest (RF) and LSTM models in distinguishing between dust and cloud classifications. The purple plot represents the spectral distribution of pixels recognized as dust by RF but identified as clouds by LSTM. Conversely, the orange plot depicts the spectral distribution of pixels identified as clouds by RF but recognized as dust by LSTM (of particular interest to us). The green and red plots correspond to the distribution of dust and cloud classifications from both models, respectively.

Interestingly, RF demonstrates superior performance in learning spectral features, as evidenced by the similar distribution of the green and purple plots (both recognized as dust by the RF model). Similarly, the red and orange plots (both identified as clouds by the RF model) exhibit a similar distribution, albeit with slight differences. The orange plot (classified as dust by LSTM) does not perfectly align with the red plot (the distribution of cloud pixels) as expected. This discrepancy between the orange and red plots highlights an interesting finding: the orange plot signifies the distribution of thick dust, with pixels exhibiting a spectral distribution similar to that of clouds. Consequently, these pixels are misclassified as clouds by the Random Forest models. This emphasizes the limitations of relying solely on spectral features for distinguishing between clouds and dust, underscoring the crucial role of temporal features learned by the LSTM model.

4.4 Discussion

Evaluation of the four models is based on their training and validation accuracy scores, which indicate their overall performance. The RF, LSTM, and RF+LSTM models show excellent performance with training accuracy scores above 0.99, and test accuracy score above 0.96. The ConvLSTM-UNet model doesn't have test dataset, but its advancement can still be observed from training accuracy score, which is more than 0.94. To further assess and compare the models, they are applied to some specific cases. The inclusion of the LSTM layer in the LSTM and RF+LSTM models enables them to capture temporal features, leading to the identification of high-albedo surfaces, clouds, and thick dust. However, the ConvLSTM-UNet model does not seem to learn temporal features as effectively. In the case of thin smoke detection, both the ABI products and the models struggle, with the ML models performing even worse than the ABI products. The ConvLSTM-UNet model show some capability to capture spatial features allows it to identify parts of thin smoke but may also result in misidentification of surface features as aerosols. Therefore, the addition of surface information to the models is necessary to improve their performance.

Objective conclusions have been derived from the statistical analysis conducted on one year of ABI observations, classified by the four ML models, across five representative areas. The comparison between the results generated by the ML models and the MODIS products underscores the robustness of the machine learning models employed. The slight discrepancies in terms of cloud fraction can be attributed to the limitations of the MODIS algorithm in effectively identifying clouds and dense aerosols. Furthermore, the highest correlation coefficient between the RF model and the MODIS products indicates the limitations of LSTM layers in accurately identifying cloud pixels and the advancement of RF to learn spectral features. Additionally, the comparison of spectral features between clouds and dust further highlights the RF model's ability to extract hyperspectral features. However, this analysis also underscores the significance of incorporating temporal features for precise identification of components.

5. Conclusion

Four machine learning models were trained to accurately identify cloud and aerosols by leveraging spectral, spatial, and temporal features. The RF model extracted spectral features, while the LSTM, RF+LSTM models utilized both spectral and temporal features, and the ConvLSTM-UNet further considers spatial features. The training datasets comprised 1 km ABI day-time observations, incorporating 9 visible and infrared channels along with 3 calculated indexes. Manual labeling was performed using NASA's Worldview web interface, and the test datasets were labeled using collocated CALIOP level 2 1 km cloud layer and 5 km aerosol layer products.

The accuracy scores and confusion matrices demonstrated that all ML models exhibited excellent overall performance, making it challenging to determine the superiority of any particular model. To further assess these models, multiple scenarios were examined, with RGB figures (1600 nm, 860 nm, and 470 nm) and ABI level-2 products (comprising aerosol detection and clear sky mask) used as reference. The results showcased the effectiveness of these models in detecting and classifying cloud and aerosol components across various scenarios. The inclusion of temporal features in the LSTM and RF+LSTM models significantly enhanced their performance. These models demonstrated proficiency in identifying high-albedo surfaces, clouds, and distinguishing between thick dust and cloud, which posed challenges for traditional algorithms. A comparison with ABI Level-2 products underscored the advantages of machine learning models over conventional approaches. Remarkably, the machine learning models displayed superior performance in diverse scenarios, including severe dust events and smoke detection.

For in-depth analysis, one-year consecutive observations from five representative areas were collected. The comparison of cloud fraction variations between the ML models and MODIS products (Terra and Aqua) demonstrates satisfactory overall performance and close correspondence. And the noticeable discrepancies observed in the west coast of Africa can be attributed to the limitations of the MODIS algorithm in accurately differentiating optically thick aerosols from clouds. Furthermore, the analysis reveals that the RF model yields the closest match with the MODIS products, which illustrates the adverse impact of LSTM layers on cloud detection. Additionally, the reliance of both the traditional MODIS algorithm and RF models solely on spectral information for pixel labeling also contribute to the high correlation coefficient and highlights the importance of incorporating temporal features for accurate component identification. The analysis of spectral features highlights the robustness of RF in extracting spectral information. However, it is emphasized that temporal variations play a crucial role in accurately classifying cloud and aerosol components. The inclusion of temporal features significantly improves the performance of the models, underscoring their importance in remote sensing applications.

Overall, this study showcases the potential of ML models for satellite-based cloud and aerosol classification. The present models exhibit excellent performance, surpassing traditional algorithms in many cases. The findings contribute to the advancement of remote sensing techniques and hold promise for enhancing weather and climate predictions. Further research is recommended to address challenges related to thin smoke detection and aerosol identification in highly contaminated regions. Future work should focus on expanding the training dataset to include more diverse scenarios, such as thin smoke, low-latitude snow/ice, and bare land without plants. Additionally, incorporating surface information into the models is crucial to improve their performance. Hybrid approaches that combine the strengths of different features and algorithms should be explored to further enhance the models' capabilities.

References

A. Taravat, F. Del Frate, C. Cornaro, and S. Vergari, "Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images," *IEEE Geosci. Remote Sensing Lett.*, vol. 12, no. 3, pp. 666–670, Mar. 2015, doi: 10.1109/LGRS.2014.2356616.

C. D. Christopoulos, S. Garimella, M. A. Zawadowicz, O. Möhler, and D. J. Cziczo, "A machine learning approach to aerosol classification for single-particle mass spectrometry," *Atmos. Meas. Tech.*, vol. 11, no. 10, pp. 5687–5699, Oct. 2018, doi: 10.5194/amt-11-5687-2018.

C. Liu *et al.*, "A Machine Learning-based Cloud Detection Algorithm for the Himawari-8 Spectral Image," *Adv. Atmos. Sci.*, vol. 39, no. 12, pp. 1994–2007, Dec. 2022, doi: 10.1007/s00376-021-0366-x.

C. Wang, S. Platnick, K. Meyer, Z. Zhang, and Y. Zhou, "A machine-learning-based cloud detection and thermodynamic-phase classification algorithm using passive spectral observations," *Atmos. Meas. Tech.*, vol. 13, no. 5, pp. 2257–2277, May 2020, doi: 10.5194/amt-13-2257-2020.

D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, "Spatiotemporal prediction of air quality based on LSTM neural network," *Alex. Eng. J.*, vol. 60, no. 2, pp. 2021–2032, Apr. 2021, doi: 10.1016/j.aej.2020.12.009.

H. Sun *et al.*, "Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement," *Geosci. Model Dev.*, vol. 15, no. 22, pp. 8439–8452, Nov. 2022, doi: 10.5194/gmd-15-8439-2022.

H. S. Kim *et al.*, "Development of a daily PM₁₀ and PM_{2.5} prediction system using a deep long short-term memory neural network model," *Atmos. Chem. Phys.*, vol. 19, no. 20, pp. 12935–12951, Oct. 2019, doi: 10.5194/acp-19-12935-2019.

I. S. Torsum and E. Kwiatkowska, "Neural network system for cloud classification from satellite images," in *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, Washington, DC, USA: IEEE, 1999, pp. 3785–3790. doi: 10.1109/IJCNN.1999.830756.

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, Aug. 2019, doi: 10.1016/j.rse.2019.03.039.

J. Lee *et al.*, "Machine Learning Based Algorithms for Global Dust Aerosol Detection from Satellite Images: Inter-Comparisons and Evaluation," *Remote Sensing*, vol. 13, no. 3, p. 456, Jan. 2021, doi: 10.3390/rs13030456.

L. A. Remer *et al.*, "The MODIS Aerosol Algorithm, Products, and Validation," *J. Atmospheric Sci.*, vol. 62, no. 4, pp. 947–973, Apr. 2005, doi: 10.1175/JAS3385.1.

L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001, doi: 10.1023/A:1010933404324.

M. A. Vaughan *et al.*, "Fully Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 10, pp. 2034–2050, Oct. 2009, doi: 10.1175/2009JTECHA1228.1.

M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.

N. Ghasemian and M. Akhoondzadeh, “Introducing two Random Forest based methods for cloud detection in remote sensing images,” *Advances in Space Research*, vol. 62, no. 2, pp. 288–303, Jul. 2018, doi: 10.1016/j.asr.2018.04.030.

O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.” arXiv, May 18, 2015. Accessed: Jan. 04, 2023. [Online]. Available: http://arxiv.org/abs/1505.04597

Q. Wang, P. M. Atkinson, and W. Shi, “Fast Subpixel Mapping Algorithms for Subpixel Resolution Change Detection,” *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 4, pp. 1692–1706, Apr. 2015, doi: 10.1109/TGRS.2014.2346535.

Q. Wu and H. Lin, “Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network,” *Sustainable Cities and Society*, vol. 50, p. 101657, Oct. 2019, doi: 10.1016/j.scs.2019.101657.

Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, “Prediction of Sea Surface Temperature Using Long Short-Term Memory,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017, doi: 10.1109/LGRS.2017.2733548.

R. A. Frey *et al.*, “Cloud Detection with MODIS. Part I: Improvements in the MODIS Cloud Mask for Collection 5,” *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 7, pp. 1057–1072, Jul. 2008, doi: 10.1175/2008JTECHA1052.1.

R. C. Levy *et al.*, “The Collection 6 MODIS aerosol products over land and ocean,” *Atmospheric Meas. Tech.*, vol. 6, no. 11, pp. 2989–3034, Nov. 2013, doi: 10.5194/amt-6-2989-2013.

R. Jing *et al.*, “Object-based change detection for VHR remote sensing images based on a Trisiamese-LSTM,” *International Journal of Remote Sensing*, vol. 41, no. 16, pp. 6209–6231, Aug. 2020, doi: 10.1080/01431161.2020.1734253.

S. A. Ackerman, R. E. Holz, R. Frey, E. W. Eloranta, B. C. Maddux, and M. McGill, “Cloud Detection with MODIS. Part II: Validation,” *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 7, pp. 1073–1086, Jul. 2008, doi: 10.1175/2007JTECHA1053.1.

S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997,doi: 10.1162/neco.1997.9.8.1735.

S. Platnick *et al.*, “The MODIS cloud products: algorithms and examples from terra,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 459–473, Feb. 2003, doi: 10.1109/TGRS.2002.808301.

T. J. Schmit, P. Griffith, M. M. Gunshor, J. M. Daniels, S. J. Goodman, and W. J. Lebair, “A Closer Look at the ABI on the GOES-R Series,” *Bulletin of the American Meteorological Society*, vol. 98, no. 4, pp. 681–698, Apr. 2017, doi: 10.1175/BAMS-D-15-00230.1.

V. Eyring *et al.*, “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization,” *Geosci. Model Dev.*, vol. 9, no. 5, pp. 1937–1958, May 2016, doi: 10.5194/gmd-9-1937-2016.

W. B. Rossow and R. A. Schiffer, “ISCCP Cloud Data Products,” *Bull. Amer. Meteor. Soc.*, vol. 72, no. 1, pp. 2–20, Jan. 1991, doi: 10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2.

W. J. Marais, R. E. Holz, J. S. Reid, and R. M. Willett, “Leveraging spatial textures, through machine learning, to identify aerosols and distinct cloud types from multispectral observations,” *Atmos. Meas. Tech.*, vol. 13, no. 10, pp. 5459–5480, Oct. 2020, doi: 10.5194/amt-13-5459-2020.

X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”.

X. Yu and D. J. Lary, “Cloud Detection Using an Ensemble of Pixel-Based Machine Learning Models Incorporating Unsupervised Classification,” *Remote Sensing*, vol. 13, no. 16, p. 3289, Aug. 2021, doi: 10.3390/rs13163289.

Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, “A CFCC-LSTM Model for Sea Surface Temperature Prediction,” *IEEE Geosci. Remote Sensing Lett.*, vol. 15, no. 2, pp. 207–211, Feb. 2018, doi: 10.1109/LGRS.2017.2780843.

Z. Zhu and C. E. Woodcock, “Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change,” *Remote Sensing of Environment*, vol. 152, pp. 217–234, Sep. 2014, doi: 10.1016/j.rse.2014.06.012.