Editorial

Statistical and Machine Learning Models for Remote Sensing Data Mining—Recent Advancements

Monidipa Das [1,*] ⬤, **Soumya K. Ghosh** [2] ⬤, **Vemuri M. Chowdary** [3], **Pabitra Mitra** [2] **and Santosh Rijal** [4]

1   Faculty in the Machine Intelligence, Indian Statistical Institute, Kolkata 700108, India
2   Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur,
    Kharagpur 721302, India; skg@cse.iitkgp.ac.in (S.K.G.); pabitra@gmail.com (P.M.)
3   Mahalanobis National Crop Forecast Centre, New Delhi 110012, India; chowdary_isro@yahoo.com
4   Department of Geography, Virginia Tech, Blacksburg, VA 24061, USA; rsantosh@vt.edu
*   Correspondence: monidipadas@hotmail.com

During the last few decades, the remarkable progress in the field of satellite remote sensing (RS) technology has enabled us to capture coarse, moderate to high-resolution earth imagery on weekly, daily, and even hourly intervals. This wealth of earth surface data, if analyzed effectively, can provide significant insights into various geo-spatial processes, and eventually, can help us in making crucial decisions in a timely manner. Nevertheless, these RS data, as continuously captured at varying spatial, spectral, and temporal resolutions, are not only voluminous but also acquired heterogeneous data, where diverse categories of sensors, i.e., optical/microwave were used. Consequently, mining useful patterns/information from these enormous volumes of heterogeneous unstructured data requires enhanced data mining techniques exploiting the power of advanced computational intelligence and high-performance computing paradigms. Moreover, in the context of resolving urgent issues, such as in environmental nowcasting, a timely analysis of the RS data requires resource-efficient computation models with real-time processing and online analysis capabilities [1,2].

With this background in mind, in this Special Issue, we called for high-quality papers focusing on recent advancements in conventional statistical as well as machine learning techniques and modern AI (artificial intelligence)-driven technologies for efficient mining of remote sensing data. This Special Issue also aimed to provide a common platform for professionals, researchers, and practitioners from heterogeneous communities, including artificial intelligence, machine learning, geographical information systems, and spatial data science, to share their views, innovations, research achievements, and solutions to foster the advancement of intelligent analytics and efficient management of remote sensing data. Papers were invited to cover the following broad topics:

- Advanced and energy-efficient machine learning models for RS data mining
- Enhanced statistical and scalable computing methods for RS data mining
- Real-time processing and online analytics of RS data
- Real-world applications of RS data mining

After the rigorous review process, a total of five papers have been accepted for publication in this issue. The selected papers either deal with the core challenges, such as missing data handling, noisy label distillation, feature-level fusion, etc., in remote sensing data analyses [3–5], or these highlight on various critical real-world problems, including oil spill detection [6], and high wind speed inversion [7].

As just mentioned above, missing data is a common problem in the field of remote sensing data analytics. This primarily occurs due to internal malfunctioning of the satellite remote sensing devices/sensors or due to the poor atmospheric condition, such as the presence of thick cloud cover. Remote sensing images with missing information not only reduce the usability of the data but also may negatively affect the performance of the

analytical models. The problem appears to be more prominent at the time of analyzing aerosol optical depth (AOD) from remotely sensed data. AOD is a key parameter reflecting the characteristics of aerosols, and plays a significant role in predicting the concentration of pollutants in the atmosphere. However, as highlighted in the work of Chi et al. [3], the AOD data obtained by satellites are often found to be missing, and thereby, impose serious research challenges. The existing methods of AOD recovery primarily focus on to the accuracy of AOD restoration while neglecting the AOD recovery ratio. In order to solve the issue, Chi et al. [3] have proposed a light gradient boosting-based two-step model, termed as TWS, that fills the missing AOD data by combining data from multiple sources and at the same time learning spatio-temporal relationships of AOD. Experimental evaluation of TWS with respect to recovering AOD from Terra Satellite's 2018 AOD product has demonstrated the reliability of TWS method in producing competitive and consistent performance in AOD restoration. Overall, the work of Chi et al. [3], as included in our Special Issue, is of great significance in the context of studying the spatial distribution of atmospheric pollutants and handling missing data in this context.

In spite of the huge availability of remotely sensed data in recent years, the data are often found to be annotated with noisy labels. Label noise occurs whenever there is a mismatch between the ground truth label and the observed label. This happens due to several reasons, including manual labeling error, wrong or misinterpretation of the data, and so on. Noisy label can lead to serious network over-fitting problem and may negatively impact on the model performance. Therefore, noisy label distillation plays an important role in remote sensing image scene classification or segmentation tasks. Traditional models are typically based on direct fine-tuning and pseudo-labeling approaches, which are not only inefficient but also, may badly influence the model in other ways. In order to address such problem, in this Special Issue, Zhang et al. [4] have proposed a novel noisy label distillation approach grounded on an end-to-end teacher-student framework, which does not require pre-training on clean or noisy data. Evaluation on benchmark remote sensing image datasets with injected noise has demonstrated the superiority of the proposed approach [4] over the state-of-the-art techniques.

Apart from dealing with the core challenges in remote sensing data analytics, another way of improving the model performance is to fully exploit the increasingly sophisticated data from multiple sources. For example, the optical remote sensing data provides us with significantly larger amount of spectral information compared to the images captured using synthetic aperture radar (SAR), whereas the SAR technology has more penetration capability and has the advantage of generating images almost in all weather conditions. Remote sensing image fusion is, thus, important for enhancing the application ability of remote-sensing images, and accordingly, it has gained immense research attention in recent years. Incidentally, the remote-sensing image fusion can be performed both at the pixel-level and at the feature-level. However, in contrast with the pixel-level fusion, feature-level fusion considers more diverse factors, and thereby, helps to obtain more macro-level information than that obtained using pixel-level fusion. This Special Issue includes an interesting article by Kong et al. [5] on feature-level fusion-based classification of remote sensing images using features extracted from polarized SAR and optical images. The approach is based on a combination of Random Forest (RF) and Conditional Random Fields (CRFs). Typically, the model exploits the power of CRF in spatial context feature modeling and improves the RF-based classification. Experimental evaluation shows the efficacy of the proposed fusion-based classification approach.

In addition to discussion on the technical challenges being faced during remote sensing image data processing, this Special Issue also includes papers on some critical applications of remote sensing data analytics [6,7]. For example, in the fourth article of this Special Issue, Almulihi et al. [6] have presented the application of SAR Image analysis in oil spill detection. The proposed approach is based on online extended variational learning of dirichlet process mixtures of Gamma distributions. The technical novelty lies here in extending the finite Gamma mixture model that can handle infinite number of mixture components. The

online learning property of the proposed model makes it more advantageous over the batch learning-based models at the time of dealing with massive and streaming data. Empirical study with respect to real-world application of oil spill detection from SAR images demonstrates the effectiveness of the approach proposed by Almulihi et al. [6].

High wind speed inversion is another critical as well as challenging application of remote sensing data analytics, which has gained significant research interest in present days. Wind speed is one of the key sea surface parameters that prominently influence diverse oceanic applications. The traditional ways of detecting wind speed using remote sensing imaging technology are often found to be failed when the wind speed is high. The study made by Zhang et al. [7], as included in this Special Issue, reveals that machine learning techniques can be effectively employed as the complements of these conventional RS technology-based models. Experimentations on multi-sourced RS data show that machine learning schemes of Support Vector Regression (SVR), combined Principal Component Analysis (PCA) and SVR (PCA-SVR), and Convolutional Neural Network (CNN) can be certainly useful for improving the accuracy in high wind speed inversion on sea surface, where CNNs are promising models in this area.

We hope that the readers will become highly benefitted from the insightful discussions and presentations of our Special Issue papers, as concisely discussed above, and also will be encouraged to contribute to these rapidly progressing areas.

References

1. Das, M. Real-time prediction of spatial raster time series: A context-aware autonomous learning model. *J. Real-Time Image Process.* **2021**, *18*, 1591–1605. [CrossRef]
2. Das, M. Online prediction of derived remote sensing image time series: An autonomous machine learning approach. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1496–1499.
3. Chi, Y.; Wu, Z.; Liao, K.; Ren, Y. Handling missing data in large-scale MODIS AOD products using a two-step model. *Remote Sens.* **2020**, *12*, 3786. [CrossRef]
4. Zhang, R.; Chen, Z.; Zhang, S.; Song, F.; Zhang, G.; Zhou, Q.; Lei, T. Remote sensing image scene classification with noisy label distillation. *Remote Sens.* **2020**, *12*, 2376. [CrossRef]
5. Kong, Y.; Yan, B.; Liu, Y.; Leung, H.; Peng, X. Feature-Level Fusion of Polarized SAR and Optical Images Based on Random Forest and Conditional Random Fields. *Remote Sens.* **2021**, *13*, 1323. [CrossRef]
6. Almulihi, A.; Alharithi, F.; Bourouis, S.; Alroobaea, R.; Pawar, Y.; Bouguila, N. Oil spill detection in SAR images using online extended variational learning of dirichlet process mixtures of gamma distributions. *Remote Sens.* **2021**, *13*, 2991. [CrossRef]
7. Zhang, Y.; Yin, J.; Yang, S.; Meng, W.; Han, Y.; Yan, Z. High Wind Speed Inversion Model of CYGNSS Sea Surface Data Based on Machine Learning. *Remote Sens.* **2021**, *13*, 3324. [CrossRef]