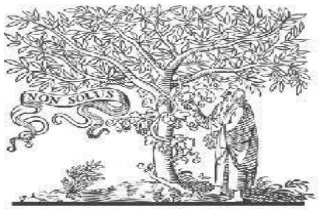
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 08th Apr 2024. Link :http://www.ijiemr.org/downloads.php?vol=Volume-13&issue=Issue4

10.48047/IJIEMR/V13/ISSUE 04/3

Title **Augmenting the Surveillance using Tailor-made Thermal Dataset for Human Activity Recognition**

Volume 13, ISSUE 4, Pages: 17-25

Paper Authors **1Sahil Mahajan,2Ajay Waghumbare,3Upasna Singh**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Augmenting the Surveillance using Tailor-made Thermal Dataset for Human Activity Recognition

[1]Sahil Mahajan, [2]Ajay Waghumbare, [3]Upasna Singh

[1,2,3]School of Computer Engg. & Mathematics Sciences
Defence Institute of Advanced Technology Pune, India
sahilmahajan1990@gmail.com, waghumbareajay@gmail.com, upasna.diat@gmail.com

Abstract
This paper presents an ameliorated approach for human activity recognition (HAR) using thermal cameras. The use of thermal imagery is crucial in scenarios where traditional cameras fail, such as night-time surveillance or challenging lighting conditions and during detection of camouflaged objects. However, there is a scarcity of open source thermal datasets due to high cost of equipment. Thus, a novel tailor-made thermal dataset DEVI (Detection and Early Warning Against Violence and Intrusion) comprising of 150,000 images is created which is designed to overcome the limitations of traditional RGB camera datasets, by incorporating multiple views, capture distances, multiple scenarios, dynamic backgrounds, illumination conditions and in-depth domain knowledge. This strategic inclusion enhances the model's adaptability and generalization across diverse environments. The DEVI also accounts for scenarios with multiple person, providing a comprehensive understanding of human interactions in crowded spaces. When tested with established models the dataset achieved high training and validation accuracy up to 98%.

Keywords — Human activity recognition , DEVI dataset ,multi views, varying illumination, dynamic background, multiple persons.

Introduction

The role of surveillance in today's world is pivotal, with applications ranging from border security to the monitoring of public spaces. These applications often demand precise HAR to ensure the safety and security of nations and communities. RGB cameras have traditionally been the workhorse of surveillance & HAR offering versatility and affordability for various applications [1]. However, their limitations become starkly evident when deployed in the complex and high-stakes scenarios like varied terrains and adverse weather conditions.

To overcome these limitations, the research focuses on harnessing modern technologies, including Machine Learning and Convolutional Neural Networks to develop a robust system that uses thermal imagery for intrusion detection and early warning [2] [3]. Thermal images have advantages such as a distribution of temperatures, making them suitable for human detection. The objective is to implement a technique for identifying human targets from thermal images, which have numerous applications in fields such as energy loss prevention, environmental monitoring, real-time surveillance and reducing false alarms. However, a significant

challenge is the scarcity of high-quality thermal datasets for training deep learning models. To address this, the paper emphasizes the creation of a customized thermal dataset DEVI, comprising approximately 150,000 diverse thermal images.

A. Limitations of RGB Cameras for HAR

RGB cameras are a ubiquitous tool for capturing visual data, primarily functioning within the visible spectrum. They have been instrumental in the development of HAR systems, facilitating the detection of activities such as walking, running, and even intricate gestures [3]. RGB cameras are versatile, offering high-resolution imagery that is suitable for many civilian applications, from traffic monitoring to crowd analysis. They are cost-effective, readily available, and can be integrated into a variety of systems, making them a common choice for video surveillance. While RGB cameras are suitable for general surveillance, they fall short in the nuanced and high- stakes field of real-life surveillance, particularly in the context of border security and various types of suspicious activity detection. In many real-life scenarios, low-light conditions are the norm. RGB cameras often struggle to capture clear imagery in low-light or nolight conditions, limiting their effectiveness in nocturnal surveillance [4]. Detecting individuals wearing camouflage clothing is a critical aspect of surveillance where RGB cameras fails drastically. RGB cameras frequently fall short in differentiating camouflage patterns from the background, complicating the identification of concealed per- sons. Lastly, real-life surveillance often involves monitoring vast, rugged terrains and urban environments. RGB cameras may struggle to distinguish between individuals engaged in different activities, such as hiding, lying prone, or exhibiting hostile intent [3] [5].

B. Limitations of Thermal Cameras for HAR

In response to these challenges, our research explores the integration of thermal cameras into surveillance systems. Thermal cameras operate in the long-wave infrared spectrum, detecting thermal radiation (heat) emitted by objects rather than visible light. The thermal videos can be recorded under various weather conditions at night, including clear, rain, and fog, and at different ranges [2]. Thermal cameras excel in low- light and no-light conditions, providing clear imaging even in complete darkness. This capability is indispensable for 24/7 surveillance. The distinct heat signatures of individuals, regardless of camouflage, make thermal cameras highly effective in identifying concealed or camouflaged targets. Thermal cam- eras can identify individuals hidden in natural environments, buildings, or trenches, simplifying threat assessment. Thus making them a potent surveillance equipment. While thermal cameras hold immense potential, there is a major missing piece in the gambit: a comprehensive thermal dataset due to high cost of cameras. Moreover, the existing thermal datasets also lacks in-depth domain knowledge.

C Motivation Behind Thermal dataset DEVI

In the modern era, border security is of utmost importance, with no country willing to tolerate even a single security lapse. To address this, surveillance cameras are being strategically installed along borders. The motivation for creating the DEVI dataset stems from the limitations observed in current RGB and thermal datasets. In the forthcoming sections, we delve into our approach to create a DEVI thermal dataset that addresses the drawbacks of existing RGB and thermal datasets in real-life

surveillance scenarios. This dataset, meticulously designed to cater to the specific requirements and classes pertinent to the surveillance domain, serves as a cornerstone for the advancement of HAR, particularly in the context of thermal imaging for border security and suspicious activity detection.

Related Work and Challenges

HAR is complex due to inter-class affinity and intra-class diversity. Despite advances in image classification methods, methods based on RGB video streams remain unsatisfactory [2] [4] [5]. Numerous efforts have been made to advance HAR utilizing data from both RGB and limited available thermal dataset coupled with a variety of Machine Learning and Deep Learning techniques. Previous research has made significant strides, but several challenges have been observed:

A. Limited Classes

The existing RGB and thermal dataset faces a crucial crunch of limited number of classes thus covering basic human activities like standing, walking, running, and lying [5]. The limitation can significantly hinder the robustness and applicability of the HAR system. A limited number of classes means that the dataset fails to encompass the full spectrum of human activities that can occur in real-world scenarios. HAR applications need to recognize a wide range of activities, from simple gestures to complex tasks, and a restricted class set may not adequately represent this diversity. HAR systems rely on machine learning models to identify patterns and make predictions. When the dataset contains a limited number of classes, the model's ability to distinguish between different activities is compromised [6] [7].

B. Lack of Views

Having a limited number of views, including capture distance, capture angles, and varied lighting conditions, in a dataset for HAR can significantly impede the robustness and real-world applicability of the trained models. This limitation hampers the ability of HAR models to recognize activities accurately in diverse and dynamic environments. With a constrained dataset, the model's capacity for generalization is compromised. It struggles to recognize the same activities when observed from different angles, distances, or light- ing conditions not present in the training data. Further, In practical applications like surveillance human activities occur under diverse conditions. A limited dataset fails to adequately prepare the model for real-world scenarios, where activities can be performed under different perspectives. Limited views result in a lack of coverage for the full spectrum of human activities. This can lead to missed or misclassified activities when the model encounters scenarios not well-represented in the training data [8] [9].

C. Based on Single Person Detection

Having a single or a limited number of persons in a dataset for HAR introduces several significant drawbacks, impacting the generalizability, diversity, and real-world applicability of the trained models. A dataset with a single or a limited number of persons inherently lacks diversity in terms of body types, clothing, and movement styles. This narrow representation makes it challenging for the model to recognize activities performed by individuals who are not part of the training dataset. The lack of diversity restricts the model's ability to adapt to different demographics and scenarios. Models trained on such datasets tend to overfit to the specific characteristics and behaviours of the individuals in the dataset.

As a result, they struggle to generalize and accurately recognize activities when applied to new, unseen individuals. This limitation undermines the model's effectiveness in real-world applications where it encounters a wide range of people. Further, such datasets also face a major challenge in scaling as they lack the requisite knowledge about diverse individuals and their activities [8].

D. Lack of Domain Knowledge

Existing RGB and thermal datasets lack domain knowledge integration, which is crucial for accurate threat interpretation. This lack of knowledge hinders models from distinguishing between normal and suspicious activities and limits their utility in training models that can understand and respond to specific nuances of border security.

By synthesizing insights from previous research, this work not only highlights the importance of thermal data but also demonstrates the significance of a diverse and specialized dataset which can work well in multiple environments [6]. The DEVI dataset aims to serve as a benchmark for future studies in HAR, providing a foundation for more accurate and adaptable surveillance systems in complex environments.

Methodology

We aim to carry out the process of creating and training our own thermal dataset DEVI (Detection and Early Warning Against Violence and Intrusion) in 2 phases.

A. Phase 1: DEVI Dataset Creation

1) Capturing Raw Videos: This was the most crucial step of the dataset creation which required a lot of deliberation since a large number of human activities were required to be captured. We shot a raw video of 800 minutes (about 13 hours) approximately with a frame rate of 30-60 fps covering various real-life aspects of HAR using a thermal camera. The thermal camera provides observation capabilities in full darkness as well as under harsh battlefield and degraded visibility conditions. These videos have been captured under different physical and environmental conditions with the aim of covering all the requirements as projected by various surveillance agencies [10] [11]. Further, we have tried to cover the variety, volume, technicality and real-life aspects as desired in various contingencies. To address the issue of limited views and perspectives each and every human activity was captured from front view, back view, side view and top view (wherever possible). For the purpose of top view, the video was shot from the terrace of 25 feet high building. These multiple views helped the dataset to adequately prepare for well-established model for real- world scenarios, where activities can be performed under different perspectives. To reduce the impact of changing lighting conditions, videos were captured during different times of the day i.e. dawn, day, dusk and night. Secondly, videos were captured using multiple modes of thermal. The modes utilized were Black Hot and White Hot. In the Black Hot mode, colder objects appear black, while hotter objects appear white or brighter. This mode is often used in situations where the focus is on detecting and highlighting warmer or hotter objects against a cooler background. In the White Hot mode, colder objects appear white or brighter, while hotter objects appear black or darker. This mode is used when the emphasis is on identifying colder objects against a warmer background. It is particularly used in scenarios such as surveillance or security, where the detection of cool or camouflaged objects against a

warmer backdrop is crucial. Further, the size of the object which is under surveillance may also vary. Sometimes the suspicious activity may be observed in the close vicinity and in some cases, it may be spotted from a greater distance. Thus, we need our system to be robust for both the contingencies [11]. So, we have recorded the raw videos in different levels of zoom varying from 1x to 3x. Thus, making the dataset more practical and robust to the situations.

It also solves the problem of concealment to a large extent. For example if a person is approaching you by hiding his weapon underneath his clothes, the heat signature of the weapon will be caught by the thermal camera, and it will be able to reproduce its shape. To make it more practical many such activities have been made the part of the dataset. To make the dataset more robust to occlusion, situations have been created in which person is either occluded behind a natural obstacles (tree, tall-grass, walls etc) or his activity is being occluded due to movement of group of men. To deliberate upon the details and to have a clear differentiation all activities were captured in slow, medium and fast motion.[12] These videos were captured in varying background as well like plains, tall-grass patches and urban areas. To add more variability and to break the shape of the silhouette, human activities were recorded in different dresses like shirt & pants, military dress, t-shirts & shorts and long kaftans.

Fig. 1. Grid of images represents various human activities being captured from multiple views. The top left image shows the side view, top right shows front view, bottom left shows back view and the bottom right shows the top view.

2) Data Pre-processing and Cleaning: In the second step, we started splitting the raw video as per different activities. Finally, we homed onto 119 distinct videos of varying length. These videos were further divided among 29 distinct activities/ classes and examined for quality. Here our aim was to improve the quality of data. Unsuitable images, such as those with low resolution, excessive noise, or heavy compression arti-facts, which significantly reduce data quality were removed. Similarly, corrupt images, which may be partially or entirely unreadable due to file corruption, are of poor quality. Data cleaning ensured that only high-quality images were included in the dataset, improving overall data quality. Secondly, the presence of corrupt images can degrade the performance of a model. Models trained on such images may struggle to extract meaningful information, leading to inaccurate predictions. Data cleaning helped to remove these problematic images, improving model accuracy and reliability. Thirdly, Clean data allows machine learning models to generalize well [10]. When unsuitable or corrupt images are present, models may struggle to generalize, as these images may contain unpredictable noise or

inconsistencies. Removing them improved the model's ability to generalize to new, unseen data. Further, we also aim to reduce bias. Bias can result from the inclusion of unsuitable images. Data cleaning also reduced bias, ensuring fair and accurate predictions. After the pre-processing and data cleaning we were able to reduce down to 85,000 images.

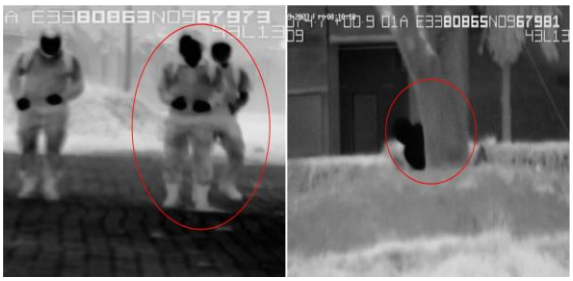


Fig. 2. Left image shows person being occluded due to movement in the group and the right image shows person hiding behind the tree getting occluded.

3) Annotation: The significance of meticulous annotation in crafting a DEVI thermal dataset is paramount, as it profoundly influences data retrieval, simplifies classification efforts, and elevates training efficacy. Annotations act as the ground truth, serving as a reference for the classification model. Accurate annotations empower the model to recognize patterns, features, and anomalies associated with distinct classes [12] [13]. Comprehensive annotations contribute to the resilience of the classification model, enabling it to associate thermal patterns with specific classes and deliver more precise predictions during testing. Labels were assigned to each image to indicate corresponding human activity. The action was carried out for 85,000 pre-processed images. Clear and com- prehensive annotations minimized the ambiguity within the dataset. This clarity is vital during the training phase, preventing confusion and ensuring the model can adeptly distinguish various thermal signatures. Well-constructed annotations also expedites the convergence of machine learning models during training. The model can efficiently adjust parameters based on accurate annotations, leading to a swifter and more effective training process. At the end of this stage we were able to divide our images into 29 classes. These many classes were created with aim of detecting maximum number of human activities required during surveillance [3].

4) Data Augmentation: In recent developments of HAR, It has been found that deep learning models are being studied by researchers, especially CNN integrated with long short term memory cells such as convolutional LSTM (ConvLSTM) networks. The deep structures require large datasets which demand extensive data collection. Therefore, various data augmentation methods under focus nowadays
[12] [13] were used. In this case, we have used techniques like crop, hue, saturation and background remove. It lead to better generalization of the data. Proper and effective augmentation further lead to efficient retrieval of images during the dataset creation process. Researchers can now easily locate, and access images based on specific criteria, streamlining the overall data fetching workflow. After this stage the total number of images available were 150,000.

5) Data Split: Dataset was split into three subsets: training, validation and test sets. The training set is used to train the model. Validation set is used for model tuning and Test set is used for evaluating the performance of the model. The ratio used is 60:20:20.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

6) Data Documentation: Firstly, a metadata file is created that contains information about each image such as file names, labels and other additional annotations. Secondly, A detailed dataset description is carried out that includes information about dataset purpose, source and other specific details about the activities. Dataset is securely stored at multiple places and backed up to prevent data loss. Further, it is password protected to prevent any unauthorized access.

Fig. 3. Left image shows a man crawling with weapon captured using 3x zoom and the right image shows similar activity in 1x zoom.

Fig. 4. Grid of images represents different human activities. Top left shows man throwing, top right shows men crossing fence, bottom left shows man lying and the bottom right shows man jumping from the wall.

TABLE I. CHARACTERISTICS OF DEVI

Characteristics	**Values**
Total Classes	29
Total Images	150,000
Data Split Ratio	60:20:20
Augmentation Techniques	Crop, Hue, Saturation, Background Remove
Models Used for Training	MobileNet, InceptionV3, Resnet50, VGG16 DenseNet201, EfficientNetB0

B. Phase 2: Training with Well-Established Models

Training accuracy indicates how well the model has learned the features present in the training data, while validation ac- curacy provides insights into the model's ability to generalize to unseen data. Training of DEVI dataset is then carried with various well-established models like mobileNet, InceptionV3, Resnet50, DenseNet201, EfficientNetB0 and VGG16 [10]. Training and validating with established models establish a baseline performance for the dataset. This baseline helps in understanding the inherent complexity and characteristics of the dataset, allowing for comparison with future models or modifications to assess improvements or challenges. By comparing the accuracy of various well- established models, we can identify which architecture performs better on the DEVI thermal dataset. This information aids in selecting the most suitable pre-trained model as a starting point for further fine-tuning or customization. The accuracy results can guide hyperparameter tuning. Adjusting parameters like learning rates, dropout rates, or layer configurations based on the accuracy metrics can optimize the model's performance on the DEVI thermal dataset.

Results and Discussion

Model	Epochs	Optimizer	Training Accuracy	Test Accuracy
MobileNet	15	Adam	89	94
IncpetionV3	15	Adam	78	80
Resnet50	15	Adam	76	82
DenseNet201	15	Adam	89	90
EfficientNetB0	15	Adam	96	97
VGG16	15	Adam	97	98

From the table II, it is evident that validation accuracy is slightly greater than training accuracy, it means that proposed model is performing well on unseen data. High accuracy testifies that DEVI successfully simulates real-world conditions, allowing pre-trained models

to adapt effectively and thus can correctly distinguish between 29 types of human activities. Further, it is suitable for transfer learning, leveraging pre- trained models' knowledge on other datasets and domains.

Conclusion

In conclusion, the creation of the DEVI thermal dataset represents a significant milestone in the field of HAR using thermal imagery. The dataset, comprising 150,000 thermal images, addresses several critical challenges faced by existing HAR datasets. These challenges included limited variety in terms of views, illumination conditions, motion, occlusion, dynamic background, variability and capture distances, all of which are essential for a robust and comprehensive HAR model. Through an extensive data collection process using a Thermal camera, we have successfully compiled a dataset that not only overcomes these challenges but also adheres to high-quality standards. The division of data into training, validation & test sets and promising results from well-established models ensures that our dataset is ready for machine learning model development. This DEVI thermal dataset trained with deep learning models opens new avenues for the research community to advance the state of the art in HAR and many related applications. Its wide variety and multi-dimensional characteristics make it a valuable resource for the development and evaluation of deep learning models and object detection techniques.

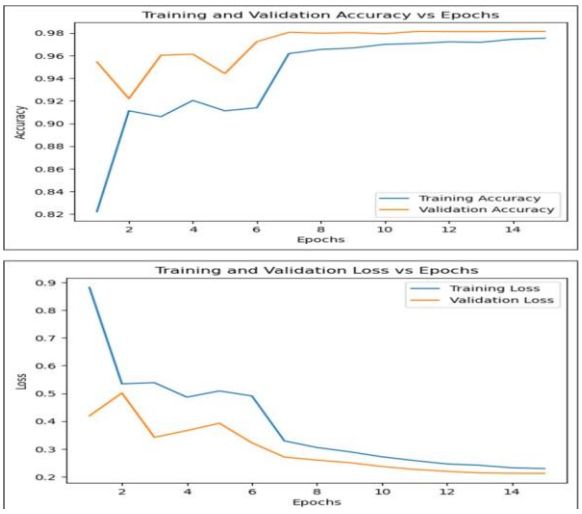


Fig. 6. Top image represents the training and validation accuracy Vs number of epochs using VGG16 for DEVI dataset and the bottom image shows training and validation loss Vs number of epochs using VGG16 for DEVI dataset

In the next phase of our research, we aim to utilize this dataset to create advanced classification models that will significantly enhance the accuracy and effectiveness of intrusion and suspicious activity detection and early warning systems. We look forward to the contributions and collaborations from the research community to harness the full potential of this DEVI thermal dataset for the benefit of society and national security. With the creation of this dataset, we are one step closer to realizing a safer and more secure world through the power of thermal imagery and deep learning.

Acknowledgment

The authors would like to acknowledge the valuable insight provided by the brave soldiers of the Indian Army who are guarding our nation from all type of

threats. Their feedback helped us in continuously improving the domain aspects of the DEVI and made it more relevant to real life scenarios.

References

[1] Poulose, Alwin, Jung Hwan Kim, and Dong Seog Han."HITHAR:Hu- man Image Threshing Machine for Human Activity Recognition Using Deep Learning Models." Computational Intelligence and Neuroscience 2022 (2022)

[2] Marina Ivaˇsi c-Kos, Mate Kriˇsto, and Miran Pobar. Human detection in thermal imaging using yolo. In Proceedings of the 2019 5th Inter- national Conference on Computer and Technology Applications, pages 20– 24, 2019

[3] Ashwani Kumar, Zuopeng Justin Zhang, and Hongbo Lyu. Object detection in real time based on improved single shot multi-box detector algorithm. EURASIP Journal on Wireless Communications and Net- working, 2020:1–18, 2020

[4] Islam, Md Milon, Sheikh Nooruddin, Fakhri Karray, and Ghulam Muhammad. "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects." Computers in Biology and Medicine (2022): 106060.

[5] Krishanu Sarker, Mohamed Masoud, Saeid Belkasim, and Shihao Ji. Towards robust human activity recognition from rgb video stream with limited labelled data. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 145–151. IEEE, 2018

[6] Pei-Fen Tsai, Chia-Hung Liao, and Shyan-Ming Yuan. Using deep learning with thermal imaging for human detection in heavy smoke scenarios. Sensors, 22(14):5351, 2022.

[7] Yu, Xianggang, Mutian Xu, Yidan Zhang, Haolin Liu, Chongjie Ye, Yushuang Wu, Zizheng Yan et al. "Mvimgnet: A large-scale dataset of multi-view images." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9150-9161. 2023.

[8] Hannes Fassold, Karlheinz Gutjahr, Anna Weber, and Roland Perko. A real-time al- gorithm for human action recognition in rgb and thermal video. In Real-Time Image Processing and Deep Learning 2023, volume 12528, pages 33–39. SPIE, 2023

[9] Waghumbare, Ajay, Upasna Singh, and Nihit Singhal. "DCNN Based Human Activity Recognition Using Micro-Doppler Signatures." In 2022 IEEE Bombay Section Signature Conference (IBSSC), pp. 1-6. IEEE, 2022.

[10] Mascarenhas, Sheldon, and Mukul Agarwal. "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Clas- sification." In 2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON), vol. 1, pp. 96-99. IEEE, 2021.

[11] More Rahul Tanaji, Detection of Human Targets from Thermal Images, International Journal Of Engineering Research Technology (IJERT) Volume 10, Issue 02 (February 2021)

[12] Bakhshayesh, Parsa Riazi, Mehdi Ejtehadi, Alireza Taheri, and Saeed Behzadipour. "The Effects of Data Augmentation Methods on the Performance of Human Activity Recognition." In 2022 8th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pp. 1-6. IEEE, 2022.

[13] Sun, Zehua, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, and Jun Liu. "Human action recognition from various data modalities: A review." IEEE transactions on pattern analysis and machine intelligence (2022).