TEMA

TRUSTED
EXTREMELY PRECISE
MAPPING AND PREDICTION
FOR EMERGENCY
MANAGEMENT

D3.2

Final report on algorithms for extreme data analytics

Funded by
the European Union

Project Information

Project acronym:	**TEMA**
Project full title:	Trusted Extremely Precise Mapping and Prediction for Emergency Management
Call identifier:	HORIZON-CL4-2022-DATA-01
Type of action:	HORIZON Research and Innovation Actions
Start date:	1 December 2022
End date:	30 November 2026
Grant agreement no:	101093003

D3.2- Final report on algorithms for extreme data analytics			
Executive Summary:	**Deliverable D3.2 Final report on algorithms for extreme data analytics, is the second deliverable of Work Package 3 (WP3) within the TEMA project. This document encapsulates the research achievements of Tasks T3.1, T3.2, and T3.3 over M19-M30 of the project. See the Section Executive Summary.**		
WP:	**3**		
Author(s):	**See table below for a full list of authors**		
Editor:	**Bernd Resch, David Hanny, Shaily Gandhi, Sebastian Schmidt (IT:U)**		
Leading Partner:	**IT:U**		
Participating Partners:	**All**		
Version:	**0.3**	Status:	**Final**
Deliverable Type:	**R Document, report**	Dissemination Level:	**Public**
Official Submission Date:	**31 May 2025**	Actual Submission Date:	**31 May 2025**

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Disclaimer

The TEMA consortium consists of the following partners:

No.	Partner Organization Name	Partner Organization Short Name	Country
1	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	GR
2	DEUTSCHES ZENTRUM FUR LUFT UND RAUMFAHRT EV	DLR	DE
3	ENGINEERING - INGEGNERIA INFORMATICA SPA	ENG	IT
4	ATOS IT SOLUTIONS AND SERVICES IBERIA SL	ATOS IT	ES
4.1	ATOS SPAIN SA	ATOS SP	ES
5	UNIVERSIDAD DE SEVILLA	USE	ES
6	TECNOSYLVA SL	TSYL	ES
7	NORTHDOCKS GMBH	ND	DE
8	INTERDISCIPLINARY TRANSFORMATION UNIVERSITY AUSTRIA	IT:U	AT
9	THE LISBON COUNCIL FOR ECONOMIC COMPETITIVENESS ASBL	LC	BE
10	LATITUDO 40 SRL	LAT40	IT
11	NELEN & SCHUURMANS TECHNOLOGY BV	NS	NL
12	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	FHHI	DE
13	UNIVERSITA DEGLI STUDI DI MESSINA	UNIME	IT
14	KAJAANIN AMMATTIKORKEAKOULU OY	KAMK	FI
16	KENTRO MELETON ASFALEIAS	KEMEA	GR
17	DIMOS MANTOUDIOU - LIMNIS - AGIAS ANNAS	D.MALIAN	GR
18	REGIONE AUTONOMA DELLA SARDEGNA	RAS	IT

This project has received funding from the European Union – European Commission under grant agreement No 101093003

19	BAYERISCHES ROTES KREUZ	BRK	DE
20	KAINUUN HYVINVOINTIALUE	KAHY	FI

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Document Revision History

Version	Description	Contributions
0.1	ToC	David Hanny, Sebastian Schmidt, Bernd Resch, Shaily Gandhi
0.2	First draft	David Hanny, Sebastian Schmidt, Bernd Resch, Shaily Gandhi, Burcu Bilgic, Heidrun Mühle, Leila Arras, Victor Prieto Ruiz, Vasileios Mygdalis, Marc Wieland, Michael Nolde, Eleonor Díaz Fragachan
0.3	Revised version	David Hanny, Sebastian Schmidt, Bernd Resch, Burcu Bilgic, Ehsan Jalilian, Leila Arras

Authors

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Name	Partner
Bernd Resch	IT:U
David Hanny	IT:U
Shaily Gandhi	IT:U
Sebastian Schmidt	IT:U
Burcu Bilgic	IT:U
Ehsan Jalilian	IT:U
Heidrun Mühle	IT:U
Leila Arras	FHHI
Victor Prieto Ruiz	DLR
Vasileios Mygdalis	AUTH
Ioanna Koroni	AUTH
Filippos Kitsos	AUTH
Nikolaos Marios Militsis	AUTH
Michael Siavrakas	AUTH
Anestis Kaimakamidis	AUTH
Polydoros Giannouris	AUTH
Dimitrios Matthaios Tzimas	AUTH
Dimitrios Papaioannou	AUTH
Evangelos Charalampakis	AUTH
Nikolaos Tzavidas	AUTH
Dimitrios Fotiou	AUTH
Marc Wieland	DLR
Michael Nolde	DLR
Eleonor Díaz Fragachan	ATOS

Reviewers

Name	Partner
Abdalraheem Ijjeh	USE
Jose Ramiro Martinez de Dios	USE

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table of Contents

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This project has received funding from the European Union – European Commission under grant agreement No 101093003

List of Figures

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This project has received funding from the European Union – European Commission under grant agreement No 101093003

List of Tables

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Abbreviations

AAG Annotator Agreement Graph.

ABEA Aspect-based Emotion Analysis.

ABSA Aspect-based Sentiment Analysis.

AEC Aspect Emotion Classification.

AI Artificial Intelligence.

API Application Programming Interface.

AR AutoRegressive.

ARIMA Autoregressive Integrated Moving Average.

ATE Aspect Term Extraction.

AttnLRP Attention-Aware Layer-wise Relevance Propagation.

BA Burnt Area.

BERT Bidirectional Encoder Representations from Transformers.

BFT Byzantine Fault-Tolerant.

BG BackGround.

CAV Concept Activation Vector.

CDSE Copernicus Data Space Ecosystem.

CE Commission Error.

CFD Computational Fluid Dynamics.

CLIP Contrastive Language-Image Pre-training.

CNN Convolutional Neural Network.

CoSy Concept Synthesis.

CRP Concept Relevance Propagation.

D-DNN Decentralized Deep Neural Network.

D3Stereo Decisive Disparity Diffusion Stereo.

DAAM Diffusion Attention Attribution Model.

DB Davies-Bouldin.

DDoS Distributed Denial of Service.

DETR DEtection TRansformer.

DINO self-DIstillation with NO labels.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

DIV-SUM DIVide and SUMmarize.

DNN Deep Neural Network.

DoA Description of the Action.

EFFIS European Forest Fire Information System.

EMS Emergency Management Service.

EnMAP Environmental Mapping and Analysis Program.

EPE End Point Error.

ESA European Space Agency.

ETM Embedded Topic Model.

FADE Feature Alignment to Description Evaluation.

FEM Finite Element Method.

FPS Frames Per Second.

FSL Few-Shot Learning.

GCN Graph Convolutional Network.

GDACS Global Disaster Alert and Coordination System.

GMM Gaussian Mixture Model.

GNN Graph Neural Network.

GPT Generative Pre-trained Transformer.

GPU Graphics Processing Unit.

GRACE GRadient hArmonized and CascadEd labeling.

GRRNN Generative-Regressing Recurrent Neural Network.

GSD Ground Sample Distance.

HSI HyperSpectral Imager.

IAMSP Individualized Agent Model Selection Process.

IID Independent and Identically Distributed.

IMDAR Invariant Multi-Descriptors for Action Recognition.

IoU Intersection over Union.

IPF Instrument Processing Facility.

IR Infrared.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

JRC Joint Research Centre.

JSTTS Joint Spatio-Temporal Topic-Sentiment.

JTS Joint Topic-Sentiment.

KD Knowledge Distillation.

KPI Key Performance Indicator.

KSA Knowledge Self-Assessment.

LAS Label Aggregation Score.

LDA Latent Dirichlet Allocation.

LFP Layer-wise Feedback Propagation.

LLM Large Language Model.

LRP Layer-wise Relevance Propagation.

LSB Localization Size Balancing.

LSTM Long Short-Term Memory.

LWIR Long-Wave Infrared.

MA Moving Average.

MACE Multi-Annotator Competence Estimation.

MAE Mean Absolute Error.

mAP mean Average Precision.

mIoU mean Intersection over Union.

MLM Masked Language Model.

MPT MosaicML Pretrained Transformer.

MSI MultiSpectral Instrument.

MWIR Mid-Wave Infrared.

NASA National Aeronautics and Space Administration.

NDM Natural Disaster Management.

NDVI Normalized Difference Vegetation Index.

NLP Natural Language Processing.

NRT Near Real-Time.

NTC Non-Time-Critical.

OE Omission Error.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

OGC Open Geospatial Consortium.

OLI Operational Land Imager.

OOD Out-of-Distribution.

PAR Privacy via Adversarial Reprogramming.

PCA Principal Component Analysis.

PCX Prototypical Concept-based Explanation.

PDE Partial Differential Equation.

PDGS Payload Data Ground Segment.

PDM Prediction and Decision-Making.

PEP Percentage of Error Pixels.

PMI Pointwise Mutual Information.

PXL-USS PIXel-based Unsupervised Semantic Segmentation.

QoI Quality of Inference.

RCNN Region-based Convolutional Neural Networks.

ReID Re-IDentification.

RGB Red Green Blue.

RMSE Root Mean Squared Error.

RNN Recurrent Neural Network.

S-BFT Shard-based Byzantine Fault Tolerant.

SAR Synthetic Aperture Radar.

SDG Stochastic Gradient Descent.

SDXL Stable Diffusion XL.

SLM Small Language Model.

SMAPE Symmetric Mean Absolute Percentage Error.

SMV Simple Majority Voting.

SOTA State of the Art.

STAC SpatioTemporal Asset Catalog.

STEGO Self-supervised Transformer with Energy-based Graph Optimization.

SU Sentiment Uniformity.

SVM Support Vector Machine.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

SVR Support Vector Regression.

SWIR Short-Wave Infrared.

TC Topic Coherence.

TD Topic Diversity.

TDA Topological Data Analysis.

TFA Trustworthy Federated Analytics.

TMV Trustworthy Majority Voting.

TQ Topic Quality.

TV Target Value.

TVSum Title-based Video Summarization.

ViT Vision Transformer.

VLM Vision Language Model.

VPSS Voting and Priority Score Selection.

VQA Visual Question Answering.

WFS Web Feature Service.

WMS Web Map Service.

WMV Weighted Majority Voting.

WP Work Package.

WRCGAN Wasserstein recurrent conditional Generative Adversarial Network.

XAI eXplainable Artificial Intelligence.

YOLO You Only Look Once.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Executive Summary

Deliverable D3.2 "Final report on algorithms for extreme data analytics" is the second deliverable of WP3 within the TEMA project. It documents the research results of tasks T3.1 "Explainable and robust analytics", T3.2 "Real-time semantic visual analysis and remote sensing" and T3.3 "Social media and text semantic analysis" carried out between M19-M30. These tasks address the challenges of extreme data analytics for Natural Disaster Management (NDM) by leveraging heterogeneous and dynamic data sources, such as satellite imagery, geospatial data, sensor inputs and multilingual social media content. WP3 focuses on the **development of novel, fast and trustworthy Artificial Intelligence (AI) algorithms** that can process complex multimodal data in real time across the edge-to-cloud continuum.

Major technical advancements include **new explainability frameworks** (T3.1) that significantly improve the speed and interpretability of local and global eXplainable Artificial Intelligence (XAI) methods, increasing trustworthiness KPIs under Objective OA1. In visual analytics (T3.2), **Deep Neural Network (DNN)-based models for fire, smoke, flood, and burnt area segmentation**, as well as **object detection**, achieved substantial gains in accuracy and speed. Notably, new methods outperformed previous state-of-the-art by up to 13.4% in classification accuracy (e.g., fire and burnt area segmentation), and up to 17% in segmentation quality for weakly-supervised tasks. Real-time analysis capabilities were confirmed across all visual components, fully addressing Objective OA3.

In the domain of text and social media analytics (T3.3), a **Joint Spatio-Temporal Topic-Sentiment (JSTTS) model** integrated multiple modalities (semantic, sentiment, spatial, temporal) and outperformed sequential baselines by up to 5× in topic quality, marking a major advancement in extreme text data analysis. Few-shot learning and spatio-temporal enrichment also led to up to 14% improvement in **relevance classification** accuracy. Additionally, a **novel sentiment annotation framework** using graph-based Trustworthy Majority Voting improved sentiment classification by up to 13.4% over the previous state-of-the-art. Furthermore, a novel method for **Aspect-based Emotion Analysis (ABEA)** was researched. These advances fulfill or exceed the KPI targets defined for objectives OA2 and OA3.

Altogether, all KPIs, objectives, and target values defined for WP3 have been successfully addressed. A total of **32 peer-reviewed publications** and **8 preprints** were produced. The developed algorithms feed into the TEMA Core and support platform components across WP4 and WP5. As a public deliverable, D3.2 not only presents key research outputs but also contributes to the dissemination of TEMAs scientific achievements in trustworthy, scalable, and rapid extreme data analytics.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

1. Introduction

1.1. Purpose and scope of the document

Deliverable D3.2 "Final report on algorithms for extreme data analytics" is the second Deliverable of the third Work Package (WP3) of the TEMA project. The main purpose of this document is to report the initial research results of Tasks T3.1 "Explainable and robust analytics", T3.2 "Real-time semantic visual analysis and remote sensing" and T3.3 "Social media and text semantic analysis" between M19-M30. Herein, this deliverable builds on the contents of Deliverable 3.1, which was submitted in June 2024.

The TEMA research efforts in the time period between M19 and M30 were focused on the following areas:

- Novel trustworthy AI algorithms for Deep Neural Networks (DNNs).
- Deep learning architectures for semantic visual analysis.
- Novel methods for satellite and Synthetic Aperture Radar (SAR) data analysis and remote sensing.
- Deep learning architectures for semantic analysis of geosocial media posts, news articles and textual content.
- Distributed semantic analysis across the edge-to-cloud continuum.

1.2. Structure of the document

NB: Each sub-section of this document contains the research progress made in the TEMA project between M19-M30 for the respective research tasks. Each sub-section 1.) briefly describes the state of the art (both internationally and the state of the research from TEMA's first reporting period M1-M18), and 2.) summarizes the research progress in the TEMA project in the second reporting period M19-M30.

The remainder of the document is structured as follows. Section 2 summarizes the main research efforts and key outputs, with respect to TEMA objectives. Section 3 lays out the development of novel trustworthy AI algorithms for DNNs. Section 4 describes the development novel semantic visual analysis and remote sensing AI algorithms for heterogeneous data modalities. Section 5 elaborates on the development of novel semantic analysis algorithms for geosocial media posts and textual content. Finally, conclusions are drawn in Section 6.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

2. Summary of the work carried out

2.1. Objectives

TEMA envisions addressing the challenges in extreme data analytics for Natural Disaster Management (NDM) like regional floods, flash floods, and wildfires by leveraging heterogeneous data sources including edge devices and sensors (e.g., drones, wind sensors, stream flow gauges etc.), satellite images, geospatial data, meteorological data, and geosocial media. These data sources are heterogeneous, voluminous, frequently updated, complex, multilingual, dispersed, sparse, and extreme in nature. The main objective of TEMA WP3 "Trustworthy Federated Analytics," (M1-M30) is to develop novel methods for accurate and fast/real-time semantic data analytics using inputs from multiple modalities. This includes creating algorithms for trustworthy AI, federated analytics, and DNN data scarcity mitigation. This involves distributed semantic analysis across the edge-to-cloud continuum to minimize latency, using AI and DNNs for trustworthy and flexible heterogeneous data analytics. Fast computations are crucial for decision-making in emergencies, supported by the vast cloud computing resources, spanning the entire edge-to-cloud continuum to meet the demands of extreme data analytics.

The specific TEMA objectives linked to Tasks T3.1, T3.2, and T3.3 are derived from existing challenges in extreme data analytics and from the socially relevant use-cases, i.e., regional floods, flash floods, and forest fires/wildfires. They are presented below, along with accompanying KPIs and TVs as defined in Section 1.1.1 of Part B of TEMAs Description of the Action (DoA). Task T3.1 contributes to TEMA objective OA1 “Increase trustworthiness of extreme data analysis algorithms”. Task T3.2 and T3.3 contribute to TEMA objectives OA2 “Increase accuracy of extreme data analysis algorithms” and OA3 “Increase responsiveness/speed of extreme data analysis algorithms”.

Section 2.2 summarizes the Research and Development (R&D) activities conducted under Tasks T3.1, T3.2, and T3.3 by TEMA partners, aligned with TEMA objectives in their natural order. First, it highlights innovations in explainable and robust AI, including new methods for scalable concept-based explainability and interpretable diffusion model outputs. Second, it outlines major advancements in visual data analytics for fire, flood, and object detection with significant gains in accuracy and speed. Satellite-based segmentation and damage assessment were enhanced through deep learning and cloud deployment. Additionally, breakthroughs in social media analysis were achieved through joint topic-sentiment modelling and multilingual classification approaches and improved sentiment/emotion analysis.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

2.2. Summary of the work carried out with respect to the objectives

Between M19 and M30, the following R&D work was carried out with respect to the objectives and KPIs of TEMA.

2.2.1. Explainable and robust analytics

Generic XAI

To address the shortcomings of state-of-the-art explainability methods, which often suffer from noise sensitivity, feature scaling issues, and lack of systematic evaluation, FHHI introduced several significant innovations. Key advancements include pattern-based Concept Activation Vector (CAV) that improve robustness and invariance, statistical learning formulation of pattern-CAV, semi-automated frameworks for spurious behavior detection and mitigation, and non-linear manifold-based concept definitions for better latent space understanding. Furthermore, FHHI proposed new training objectives for disentangled concept steering, automatic evaluation frameworks for textual feature descriptions (Concept Synthesis (CoSy), Feature Alignment to Description Evaluation (FADE)), XAI-based pruning and gradient-free training methods (Layer-wise Feedback Propagation), and released the open-source library quanda for systematic benchmarking of training data attribution methods. These efforts advance both the interpretability and robustness of AI models, exceeding the "**Speed of local XAI**" and "**Speed of global XAI**" KPIs of Objective **OA1**, by providing fast, scalable, and quantifiable explanation generation and evaluation tools across multiple modalities, ensuring that local and global explanations are delivered in near-real-time.

XAI on diffusion models for image generation

To overcome the opacity of diffusion model generation processes and improve transparency in automated image labelling, ATOS adopted advanced explainability tools such as Attention Map Diffusers. These methods extend beyond prior work like Diffusion Attention Attribution Model (DAAM) [21] by enabling precise token-level attention mapping during the de-noising process in state-of-the-art diffusion models including Stable Diffusion XL, SD 3.5, and Flux.1. ATOS leveraged these capabilities to develop scalable, interpretable image annotation pipelines where generated attention maps are used to automatically derive bounding boxes for specific prompt elements, such as "fire" or "smoke" in drone imagery. This advancement enhances both the interpretability and operational usability of diffusion-generated imagery, substantially accelerating the labelling process while maintaining semantic accuracy. The fast, token-level explainability addresses the "**Speed of local XAI**" KPI under Objective **OA1**, ensuring that explanations are produced efficiently enough to be integrated into near-real-time data pipelines. Details of these developments are presented in Section 3.3 and demonstrated in Figures 4 to 8.

Decentralised DNN inference for forest fire classification

To overcome the scalability and robustness limitations of previous decentralized inference methods such as Majority Voting and Quality of Inference (QoI) [22], described in D3.1, AUTH introduced the Shard-based Byzantine Fault Tolerant (S-BFT) consensus mechanism. This novel

This project has received funding from the European Union – European Commission under grant agreement No 101093003

approach partitions the decentralized DNN network into task-relevant shards using Out-of-Distribution (OOD) detection, enabling consensus only among nodes with domain expertise. The S-BFT protocol significantly improves classification accuracy, outperforming previous methods like Majority Voting and QoI by over **20**%20 on fire classification and relevance classification datasets.

The systems resilience to Byzantine faults, combined with its efficient scalability, marks a major advancement in real-time decision-making for high-risk environments. These improvements exceed the **"Image Recognition Accuracy"** KPI of Objective **OA2** by providing a robust and scalable inference framework. Additionally, the method preserves decentralized, real-time operation, directly supporting the objective **OA3** to increase responsiveness. Full technical details are provided in Section 3.3 and a related conference publication [23].

2.2.2. AI algorithms for visual data analysis

Fire and smoke region segmentation

To address the limitations of single-modality fire and smoke detection systems, AUTH developed a novel intermediate image fusion architecture that combines visible (RGB) and Infrared (IR) modalities using dynamic attention mechanisms within a lightweight real-time DNN. This architecture integrates a U-Net-inspired decoder with skip connections to enhance spatial precision, enabling robust segmentation even in the presence of dense smoke, sensor misalignment, and partial sensor failures. Evaluated on the FLAME2 dataset [1], the proposed method outperforms all SOTA fusion-based segmentation models across all classes and metricsachieving a 94.37% average recall and 88.17% mIoU, with notable improvements in the challenging fire class. These results clearly demonstrate a significant advancement in both **Segmentation accuracy** (Objective **OA2**) and inference speed (Objective **OA3**), satisfying the **Visual analysis speed** KPI for real-time deployment in high-risk wildfire scenarios. This method is described in Section 4.2.1 and a technical report that has been submitted in the form of a journal paper.

To tackle the scarcity of annotated data in fire region segmentation, AUTH proposed PIXel-based Unsupervised Semantic Segmentation (PXL-USS), a novel method that leverages sparse pixel-level prompts and dynamic thresholding for unsupervised learning. By using minimal annotations as class prototypes and dynamically computing contrastive loss thresholds within the self-DIstillation with NO labels (DINO) [24] feature space, the method ensures effective feature clustering without manual hyperparameter tuning. Evaluated on the Corsican Fire Database [14], PXL-USS achieves up to 78.23% mIoU, outperforming previous SOTA unsupervised models like Self-supervised Transformer with Energy-based Graph Optimization (STEGO) by **17**%, while requiring significantly fewer computational resources. This surpasses the **Semantic segmentation accuracy** KPI under Objective **OA2**, validating its accuracy and practicality in resource-constrained, real-time fire detection applications. This method is described in Section 4.2.2 and a technical report that has been submitted in the form of a journal paper.

Flood region segmentation

To address the visual appearance imbalance between flooded regions (foreground) and cluttered backgrounds in real-world imagery, AUTH developed a novel Self-Knowledge Distillation (KD) framework. This method trains a Teacher model on variance-reduced, augmented inputs while the Student model learns from real data, enabling it to focus on flood-discriminative features and suppress background noise. The approach is model-agnostic, operates exclusively during

This project has received funding from the European Union – European Commission under grant agreement No 101093003

training, and introduces no inference-time overhead, making it ideal for real-time flood segmentation. Applied to PIDNet [6] and other real-time semantic segmentation models, the method consistently improves mean Intersection over Union (mIoU) performance over both baseline and feature-based KD variants. In particular, it achieves a **1**% absolute mIoU improvement over AUTH's previous SOTA reported in D3.1 and a **4.5**% gain compared to prior methods on the FloodSeg dataset [25]. These results fulfill Objective **OA2** by significantly "Increasing accuracy in extreme data analysis". Furthermore, as the method maintains real-time processing capability, it directly contributes to Objective **OA3**, achieving the "**Visual Analysis Speed**" KPI. This method is described in Section 4.2.5 and a conference paper [26].

Person and car detection in flooded areas

To address training instabilities in object detection tasks involving significant variation in object sizes, AUTH introduced a coordinate-based weighting strategy for the L_1 loss, referred to as Localization Size Balancing (LSB), which emphasizes smaller objects during training. This method was integrated into RT-DETR-R18 [8], a real-time DEtection TRansformer (DETR)-based detection model, and evaluated on challenging scenarios such as person and car detection in flooded regions, as well as on the VisDrone dataset [18], which contains small and densely packed persons and cars. The results demonstrate consistent improvements in both mean Average Precision (mAP) and mAP50 metrics compared to YOLO-based baselines. This satisfies Objective **OA2** by increasing the accuracy of extreme data analysis algorithms by **5**%, particularly in complex environments. Additionally, due to its integration in RT-DETR, a real-time detection framework, AUTH method meets Objective **OA3**, improving responsiveness and exceeding the KPI for "**Visual analysis speed**" in time-critical applications. This method is described in Section 4.2.7 and a conference paper [27].

Visual privacy preservation

Through the development of two novel, privacy-preserving systems, one for facial anonymization based on the lightweight CenterFace model and another for license plate anonymization using YOLOv11-small, AUTH has demonstrated high-speed, resource-efficient visual analysis solutions suitable for real-world deployment. Both systems operate in real time, even on low-power edge devices, and support batch processing, enabling rapid analysis in large-scale surveillance and emergency response scenarios. The facial detection method ensures privacy in crowded or sensitive environments such as shelters, while the license plate detection system enables privacy-aware monitoring in transportation and public spaces. Together, these methods exemplify how extreme-scale visual data can be processed quickly and securely, without compromising individual privacyfully addressing the responsiveness and efficiency goals set out in **OA3** "Increase Responsiveness/Speed of Extreme Data Analysis Algorithms." The developed method is described in detail in Section 4.2.8.

2.2.3. Satellite data analysis

Satellite-based flood detection and assessment

To further advance the satellite-based flood detection and assessment technology, DLR extended its modular processing system by integrating new DNN models optimized for water segmentation across multiple satellite sensors and deploying the automated processing chain on a

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Kubernetes cluster. This enabled faster processing and result dissemination through publicly accessible Open Geospatial Consortium (OGC) Web Map Service (WMS) and SpatioTemporal Asset Catalog (STAC) Application Programming Interface (API), allowing seamless integration into the TEMA platform. Additionally, DLR enhanced the reference datasets and demonstrated the added value of fusing remote sensing data with social media information for rapid hotspot identification in disaster scenarios. Major speed improvements were achieved through the direct downlink of Sentinel-1 data at the DLR Neustrelitz receiving station, reducing the time between image acquisition and availability for analysis by a factor of **5** compared to the Copernicus Data Space Ecosystem. Moreover, DLR developed and benchmarked object detection models on very high-resolution (<1m Ground Sample Distance (GSD)) orthoimages, achieving a **mAP@0.5 of 0.57** for buildings and vehicles with YOLOv5l models, ensuring robust cross-platform generalization critical for disaster response. These advancements significantly exceed the TEMA KPIs for Objective **OA2** "Increase accuracy of extreme data analysis algorithms" and Objective **OA3** "Increase responsiveness/speed of extreme data analysis algorithms," by improving flood segmentation accuracy, dramatically reducing end-to-end data processing times, and enabling faster, higher-frequency disaster monitoring. Real-world deployments during the Thessaly floods (Greece, 2023), the Southern Germany floods (2024), and the Central Europe floods (2024) validated the operational readiness and high scalability of the system implemented within TEMA, with over 1,000,000 km^2 monitored and hundreds of real-time flood masks generated within days,

Satellite-based fire detection and assessment

DLR has enhanced its Burnt Area (BA) detection system by developing a Near Real-Time (NRT) deep learning-based mapping framework that combines a novel superpixel-based segmentation technique with a Graph Convolutional Network (GCN) model. Building on the flexible multi-sensor processing chain, the new DLRBAv2NRT system achieves daily BA perimeter generation using mid-resolution optical imagery and active fire data. Extensive evaluations against previous DLR rule-based methods (DLRBAv1NTC [28]) and established machine learning models (Random Forest, LightGBM), as well as comparisons with standard products such as MCD64A1v061 and CGLBA31nrt, demonstrate major accuracy and timeliness gains. Specifically, DLRBAv2NRT achieved an average Intersection over Union (IoU) of **0.69** and F1-score of **0.81** with a product availability of under one houroutperforming CGLBA31nrt (IoU 0.62) and MCD64A1v061 (IoU 0.67). The monthly-refined DLRBAv2NTC product reached an even higher IoU of **0.71**. These improvements result in a **7% IoU gain** over previous standard products and build on earlier TEMA developments that already demonstrated a **23% improvement** over Copernicus Emergency Management Service (EMS), JRC European Forest Fire Information System (EFFIS), and NASA MCD64A1 baselines. These advancements substantially exceed the TEMA KPIs for Objective **OA2** "Increase accuracy of extreme data analysis algorithms" and Objective **OA3** "Increase responsiveness/speed of extreme data analysis algorithms," by providing more accurate and faster burned area mapping capabilities suitable for operational emergency response and environmental monitoring.

2.2.4. Social media and text semantic analysis

Semantic topic modelling

Within TEMA, IT:U achieved major advancements beyond the state-of-the-art in semantic analysis of social media data, directly contributing to **OA2 "Increase accuracy of extreme data analysis algorithms"** and **OA3 "Increase responsiveness/speed of extreme data analysis al-**

This project has received funding from the European Union – European Commission under grant agreement No 101093003

gorithms". An improved Joint Spatio-Temporal Topic-Sentiment (JSTTS) model was developed during the second reporting period, successfully integrating semantic, sentiment, spatial and temporal information into a unified framework. Across four multilingual disaster datasets, JSTTS consistently outperformed traditional sequential workflows, achieving up to $5\times$ higher Topic Quality (TQ) scores and significantly improving Sentiment Uniformity (SU) (+29%) within clusters. In addition, a topic summarisation and information extraction module based on Llama-2 was developed, further improving interpretability for rapid disaster response applications. Multilingual processing capabilities were systematically validated, showing robust performance across English, German, and Spanish datasets. Parallel to this, IT:U advanced relevance classification by developing a few-shot learning framework using small language models (e.g., GPT-4o-mini, Qwen2.5 7B) and contrastive learning with SetFit, achieving an average macro F1-score of 0.77 and outperforming previous methods by up to 14%. The integration of spatio-temporal features with TwHIN-BERT-base further boosted performance to a macro F1 of 0.814. These advances demonstrate substantial gains in semantic analysis speed and accuracy, allowing real-time or near-real-time extraction of actionable information from multilingual social media streams, and fully meeting the targeted KPIs for extreme data analysis under OA2 and OA3.

Text sentiment analysis

Within TEMA, substantial advances were achieved in sentiment and emotion analysis of short texts, directly contributing to **OA2 "Increase accuracy of extreme data analysis algorithms"**. AUTH developed the novel graph-based Trustworthy Majority Voting (TMV) method, which improved annotation robustness and achieved a **balanced accuracy increase of up to 13.4%** compared to the previous state-of-the-art Loss-Modeling approach. Furthermore, the AnchorBERT model was introduced for multi-label sentiment classification, combining the advantages of independent binary classification and sentiment interrelation modelling, leading to consistently higher classification performance across several datasets. In emotion analysis, the adaptation and fine-tuning of the GRadient hArmonized and CascadEd labeling (GRACE) model for aspect-based emotion analysis (EmoGRACE) resulted in a **6 percentage point** F1-score improvement, establishing a new baseline for fine-grained emotional content detection. High-quality training datasets were created through consensus-based labelling, further strengthening model reliability. These advancements demonstrate major relative improvements in both sentiment and emotion classification accuracyexceeding the **+10% accuracy improvement** threshold targeted by **OA2**. In addition, the focus on lightweight model architectures and improved annotation workflows contributes to **OA3 "Increase responsiveness/speed of extreme data analysis algorithms"**, by enabling faster, more robust data processing and model retraining.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

3. Explainable and robust analytics

3.1. Introduction

The TEMA system heavily relies on AI/DNN-based data analysis to make predictions on multiple data sources and modalities (for instance, using drone images to detect people or segment flood areas, or using textual data from social media posts to identify relevant information from disasters). Beyond the mere predictions' quality, it is also essential, especially in such a high-stake scenario like natural disaster management, to **uncover the models' decision strategies and ensure robustness of the predictions**, to enable trust and reliability in the AI system. This is the focus of the present **Task T3.1 "Explainable and robust analytics"**.

To achieve this objective, TEMA investigates **new XAI algorithms** for various input modalities and architectures, both in terms of explaining **individual predictions** (i.e., local explanations), but also in terms of **global decision strategies** using concepts and prototypical samples (i.e., global explanations). Global explanations allow the detection of **spurious model behaviors** and encompass **solutions to mitigate** such issues, as well as the **identification of outlier** (Out-of-Distribution (OOD)) samples, hence making model decisions more **robust**.

Alongside meeting these objectives, the **faithfulness** of explanations shall be evaluated through **quantifiable metrics**, and the **computation times** of explanations (both local and global) should meet specific **speed requirements** (KPIs) to allow deployment of XAI in **near-real-time**, both are also addressed through TEMA advancements, in order to fulfill the TEMA **Objective OA1 "Increase trustworthiness of extreme data analysis algorithms"**.

FHHI is leading the present Task and focuses on **generic XAI** methods. Parts of the Task output are integrated into the TEMA software platform and used as input to **Task T5.3 "Augmented Reality and rapid visualization"**, as detailed for instance at the end of Section 3.2. **ATOS** focuses on **XAI for Diffusion Models in image generation**, while **AUTH** develops methods for **AI robustness**. **UNIME** provides links to **Task T3.4 "Real-time federated analytics"** by developing the edge-to-cloud computational infrastructure to run the **XAI component (TFA-tech-02)** of TEMA.

3.2. Generic XAI

SOTA (incl. TEMA M1-M18)

XAI has become an essential research field for ensuring transparency and trustworthiness in machine learning models, particularly DNNs. **Generic XAI** approaches span a wide variety of data modalities, but most prominently focus on **classification tasks** in computer vision and natural language processing, and very few works tackle other tasks such as **object detection or segmentation**. Further, there is a **lack of explainable DNN models for fire and flood detection** in the NDM context. Both are research gaps that TEMA aims to close.

Common **post-hoc explanation techniques** such as Layer-wise Relevance Propagation (LRP) [29], Occlusion [30], and Gradient Saliency [31] aim to highlight input features responsible for model predictions. However, the **evaluation of XAI methods** remains a challenge due to the lack of ground truth explanations in most scenarios, resulting in conflicting quality estimations.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Recent TEMA advances such as MetaQuantus [32] propose a **meta-evaluation framework to benchmark XAI quality estimators**, this framework was released by FHHI under an open-source license. A widely-used technique to sanity check explanations without ground truths is to perform **randomisation tests at the model or data level**. TEMA work [33] extended the former by proposing to minimize the effect that noise has on the evaluation through sampling, and by replacing potentially biased pairwise similarity metrics with an explanation complexity measure (based on discrete entropy). Lastly, another evaluation approach for XAI is to compare the **usefulness of explainability methods from an end-user perspective** [34], which has also been addressed through previous FHHI work.

Meanwhile, global explanation techniques have emerged, providing concept-based insights into the model's decision strategies, in particular through **Concept Activation Vectors (CAVs)** [35]. Such explanations have applications for **model correction**, i.e. for mitigating model reliance on spurious correlations, although such mitigation must additionally manage potential overcorrection to avoid degrading prediction performance. The **reactive model correction approach R-ClArC**, building on P-ClArC, was proposed by FHHI to counteract overcorrection based on model-derived knowledge and XAI insights [36]. To enhance concept-level corrections, **gradient-based penalization** was introduced to robustly mitigate model biases using CAVs [37], and the robustness of CAVs was further improved beyond traditional Support Vector Machine (SVM)-based directions.

Prototypical Concept-based Explanations (PCXs) leveraging latent relevance distributions using Gaussian Mixture Model clustering on top of concept-based relevances enabled new approaches to **outlier detection** [38]. More precisely, the latter XAI method developed by FHHI enables to **quantify the similarity between a new prediction and a prototypical model prediction** (i.e., a cluster centroid), alongside **inspecting the semantic of concepts** through visualization of **relevance-maximizing reference samples**. In the previous SOTA, PCX was applied **solely to image classification models** (using ResNet, VGGNet, EfficientNet architectures) trained **on standard and widely-used image classification tasks** (such as ImageNet, CIFAR-10 and CUB-200) [38]. During the first reporting period of TEMA (M1-M18), FHHI extended the PCX method **for the first time to a natural disaster management use-case** by applying this global explanation technique to a **fire classification network** using an EfficientNet-B0 architecture that was trained on a **publicly available fire classification dataset** [39], as was already reported in the Deliverable D3.1. During the current reporting period of TEMA (M19-M30), FHHI extended the PCX method **beyond classification models**, by **widening its scope of application for the first time to both localization and segmentation models**. This was achieved by leveraging and combining the PCX method, which was so far only applied to classification models, with Concept Relevance Propagation (CRP) for Localization models (L-CRP) [40], a SOTA XAI method providing concept-conditional attributions on localization and segmentation models, where concepts correspond to channels (i.e., filters) in a convolutional layer of the network. In particular, FHHI **applied both PCX and L-CRP to various NDM use-cases**: a **YOLOv6s6 person and car localization model developed by AUTH**, as well as **multiple UNet models for flood and fire segmentation developed by DLR and AUTH** resp. in the context of the TEMA WP3. More details on these models and tasks can be found in the following Subsection "Advances beyond SOTA: M19-M30".

Besides, FHHI work on mechanistic interpretability led to methods for **disentangling polysemantic neurons** [41]. The **interpretability of predictive uncertainty** was explored in [42], for the particular case where the uncertainty is computed as a variance over an ensemble of predictions.

Beyond computer vision, FHHI **extended XAI to time series data**: DFT-LRP [43], a method that combines the discrete Fourier transform and Layer-wise Relevance Propagation, enabled

This project has received funding from the European Union – European Commission under grant agreement No 101093003

interpretable frequency-domain analyses. Audible heatmaps for audio classification as well as the open-source audio dataset AudioMNIST were introduced in [44], and comparative analyses of input representations (waveform vs. spectrogram) using XAI were conducted to uncover model decision strategies for audio event detection [45]. Another study revealed concepts on an MRI-based classification model [46].

A further work by FHHI on generic XAI methods is **DualView** [47], a new **scalable training data attribution** method, where the goal is to find training datapoints that are responsible for a given prediction.

Advances beyond SOTA: M19-M30

In the current reporting period (M19-M30), FHHI extended its **research on generic XAI methods** by introducing new methods that address key challenges such as noise robustness of CAV directions, statistical interpretability, and concept disentanglement.

Robust Concept Directions. In [48] FHHI proposed **pattern-based CAVs**, a new method for concept direction estimation in DNNs which focuses solely on the concept's signal while filtering out distractor features. This results in concept directions which are **invariant to feature scaling and more robust to noise rotation** in the latent space, improving the reliability of both concept-sensitivity testing and model correction applications. Building upon this, a subsequent study by FHHI [49] introduced a **statistical learning formulation of pattern-based CAVs using ridge regression**, allowing the **mean and covariance of concept directions to be derived analytically**. This enables **more stable and statistically interpretable concept representations**, especially in cases where **uncertainty estimation** is critical. Together, these two contributions advance the robustness, precision, and reliability of concept-based interpretability in deep learning models.

Model Correction and Concept Alignment. In [50] FHHI proposes a **semi-automated XAI-based framework** for the **detection and mitigation of spurious shortcut behaviors** in computer vision DNNs using Concept Activation Vectors learned on top of expert-validated biased samples. In [51] FHHI proposes a **novel concept definition as non-linear manifold** and leverages a generalized Rand index with pseudo-metric properties to measure **concept-based alignment**, providing **insights into the structure of latent representation** within and across Vision Transformer (ViT) models. Lastly, in [52] FHHI proposes a novel **CAV-based training objective penalizing non-orthogonality between concept directions**, thereby **encouraging disentangled representations** in latent space with a positive impact on CAV-based steering tasks, i.e., **allowing one to add or remove concepts in isolation without impacting correlated concepts**.

Evaluating Textual Feature Descriptions. Another line of research explored by FHHI concerns **explanations of features** within DNNs delivered in the form of **open-vocabulary textual descriptions**, and in particular how to **automatically evaluate such explanations**. The **Concept Synthesis (CoSy)** framework proposed by FHHI in [53] is an **automatic framework for evaluating textual explanations of neurons in Computer Vision models** that leverages a text-to-image generative model to generate synthetic images from descriptions which are used for inference to collect activations, and then compare them to those collected on control images, in order to **quantitatively assess the quality of the textual explanations**. The **Feature Alignment to Description Evaluation (FADE)** framework [54] proposed by FHHI is another such automatic evaluation framework, but focusing on **descriptions of features in Large Language Models (LLMs)**. It **leverages an evaluating LLM** and a natural dataset to automate the evaluation process, in particular the evaluating LLM is used for rating the strength of concept expression in

This project has received funding from the European Union – European Commission under grant agreement No 101093003

samples and for generating synthetic concept data.

Evaluating Local Attributions. Since numerous **local explanation methods** exist in the literature that can explain a model's decision, either by **assigning attribution scores to input features (i.e., feature attribution methods)**, or by **assigning scores to training data samples (i.e., training data attribution (TDA) methods)**, a **standardized and systematic evaluation** of attributions is required to make **quantifiable** progress in XAI. Hence FHHI releases an **open-source Python library (named quanda) for benchmarking TDA methods** [55] under an MIT Licence. It includes various metrics, datasets and a unified interface for major existing TDA methods, with a main focus on the Computer Vision domain. Similarly, in order to **systematically evaluate feature attribution methods on Transformer LLMs**, FHHI proposes an evaluation study for various decomposition-based XAI methods on Transformers [56] together with the **release of carefully constructed ground truth annotations** for benchmarking attributions on language models. The latter work by FHHI [56] also introduces a **faster and simpler implementation of the recently proposed Attention-Aware Layer-wise Relevance Propagation (AttnLRP) method** for Transformers [57], reaching a **speedup of 1.5x in computation time over the previous SOTA**.

XAI for Pruning and Gradient-free Training. Besides local and global explanations, FHHI investigated additional applications of explanations such as **improving model efficiency and training**. In [58] FHHI proposes to explicitly **optimize the hyperparameters of attribution methods (i.e., of local explanations) to prune transformer and convolutional vision models**, achieving **higher sparsification rates** while still maintaining high prediction performance on classification tasks. In [59] FHHI introduces **Layer-wise Feedback Propagation (LFP)**, a **novel XAI-based gradient-free training** procedure which has applications for neural network pruning, as well as for the **training of non-differentiable models such as Spiking Neural Networks**.

Beyond the previously mentioned **research outcomes on generic XAI methods**, that span different input modalities (such as images, text, time series), and cover both concept-based as well as local attribution methods, including the quantitative evaluation and benchmarking of such methods, FHHI also **extended various XAI methods to the TEMA-specific needs in natural disaster management**, in particular by **applying XAI on AI models developed by other TEMA partners for various use-case scenarios of WP3**. More concretely, during the current reporting period (M19-M30), FHHI achieved the following **tailored XAI developments** through fruitful collaboration with other TEMA partners:

- **Extension and application of the global XAI method PCX** [38] introduced by FHHI during the first reporting period (M1-M18) to **localization and segmentation AI models developed by AUTH on drone images** for WP3. The PCX method is a concept-based XAI method based on a Gaussian Mixture Model (GMM) clustering of concept-based relevances, and it allows the identification of **global model decision strategies through prototypes** (i.e., cluster centroids), as well as the **quantification of similarity between a new prediction and prototypical predictions**, thus enabling the detection of outliers. In the previous TEMA SOTA, the PCX method was **solely applied to classification models**, a novel TEMA development during M19-M30 is the **extension of this method to localization and segmentation models**. More precisely, FHHI implemented PCX on a **YOLOv6s6 localization model for the detection of Persons and Cars** in a NDM context (for instance in flooded regions), as well as on a **U-Net flood segmentation model**. Both AI models were developed by AUTH within **Task T3.2 "Real-time semantic visual analysis and remote sensing"** of WP3, more particularly within the context of the Subtask "Person and car detection in flooded areas" described in Section 4.2.7, as well as the Subtask "Flood region segmentation" described in Section 4.2.5 of the present Deliverable D3.2. The PCX implementation was

This project has received funding from the European Union – European Commission under grant agreement No 101093003

also **integrated into the TEMA software platform as the XAI component "TFA-tech-02"**, and **deployed onto the TEMA cluster**, in order to enable end-users and technical partners to access concept-based XAI visualizations of the models' predictions in near-real-time.

- **Application of the global XAI method L-CRP** introduced in [40] by FHHI to **segmentation AI models developed by DLR on satellite images** for WP3. **CRP [60] is a concept-based XAI method**, where **contributions to the predictions are assigned to latent features**, mainly channels in hidden convolutional layers (each channel encoding a specific concept). **L-CRP is an extension of CRP to localization and segmentation models**, it allows to deliver concept-conditional heatmaps, and to identify potential model biases (like reliance on context instead of object features). More precisely, FHHI implemented L-CRP on **two U-Net based segmentation models for the identification of flood and burnt area**. Both AI models were developed by DLR and trained on Sentinel-2 satellite images within **Task T3.2 "Real-time semantic visual analysis and remote sensing"** of WP3, more particularly within the Subtasks "Satellite-based flood detection and assessment" described in Section 4.3.1 and "Satellite-based fire detection and assessment" described in Section 4.3.2.
- **Application of the local XAI method AttnLRP** introduced by FHHI in [57] (and which made more efficient in follow-up study [56]) to **a Transformer-based AI model developed by ITU for the classification of social media posts into disaster-relevant and disaster-irrelevant information** for WP3. AttnLRP is a SOTA decomposition-based attribution method, that assigns a contribution towards the prediction to single neurons, both in the input layer, as well as in hidden layers. Such contributions can be summed-up across embedding dimensions to obtain **token-level contributions**, and can be further used to **build document embeddings** representing the entire document. This allows one to **trace back the model's classification decision to the most relevant tokens** per document or category, as well as to **explore the semantic embedding space of documents**. More precisely, FHHI implemented AttnLRP on a **multilingual XLM-RoBERTa classification model for disaster relatedness classification**, this AI model was developed by ITU within **Task T3.3 "Social media and text semantic analysis"** of WP3. For further details on this collaboration between FHHI and ITU we refer to Section 5.6.

In the Figures 1 and 2 we provide some exemplary results fo PCX on the YOLOv6s6 and U-Net models from AUTH. In these XAI visualizations the top left image represents the current prediction, and the top right image is the nearest prototype from the dataset. Each row corresponds to one top concept (i.e., a channel in a given convolutional layer of the model), and in the middle column we visualize the concepts via retrieving the relevance-maximizing samples from the dataset for each concept. Most importantly the "Difference to prototype" column quantifies the difference in concept usage between the actual prediction and the prototypical model's decision. In Figure 1 we see a prediction for flood segmentation that is similar to a prototypical decision of the model for all 3 concepts. Although there is some redundancy in the concepts (meaning all concepts include flood in vegetation areas), the top concept (number 173) is more focused on roads and bridges, while the second and third concepts (85 and 174) further contain habitations. In Figure 2 we see a prediction for the localization of a white car. Here the top concept (number 49) represents a windshield, and this concept is over-used in the actual prediction compared to the prototype. The second and third concepts (121 and 94) represent a vehicle corner and hood.

In the Figure 3 we visualize all seven prototypes identified by PCX (top images) for the YOLOv6s6 car localization model from AUTH, together with the set of most relevant concepts (left images) alongside their concept relevance scores (numbers in the grid). We observe that the model has learned to identify different prototypes of cars: red ambulances (prototype 3), white trucks (prototypes 2 and 5), as well as ordinary cars of different colors (prototypes 0, 1, 6, 7). The

This project has received funding from the European Union – European Commission under grant agreement No 101093003

top used concepts are vehicle parts with the same colors as the prototype. Indeed the red concept (number 41) is almost exclusively used to detect red ambulances (prototype 3), the blue concept (33) is mainly used to detect cars of the same color (prototype 7). This comprehensively illustrates the different global decision strategies used by the model to identify cars.

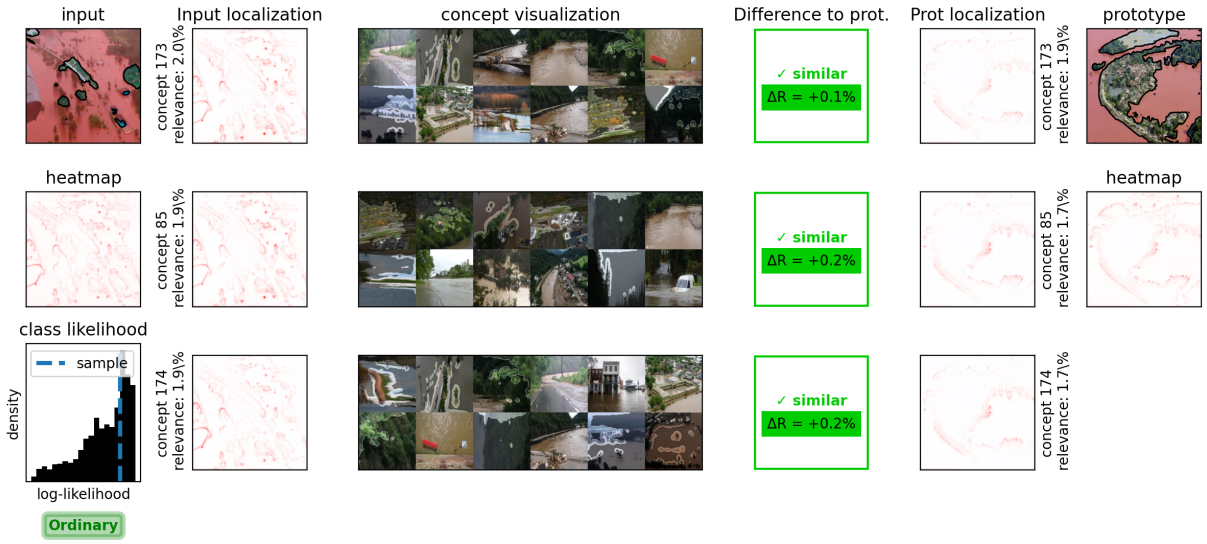


Figure 1. Example PCX visualization on a flood segmentation prediction using the U-Net model from AUTH.

Figure 2. Example PCX visualization on a car localization prediction using the YOLOv6s6 model from AUTH.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

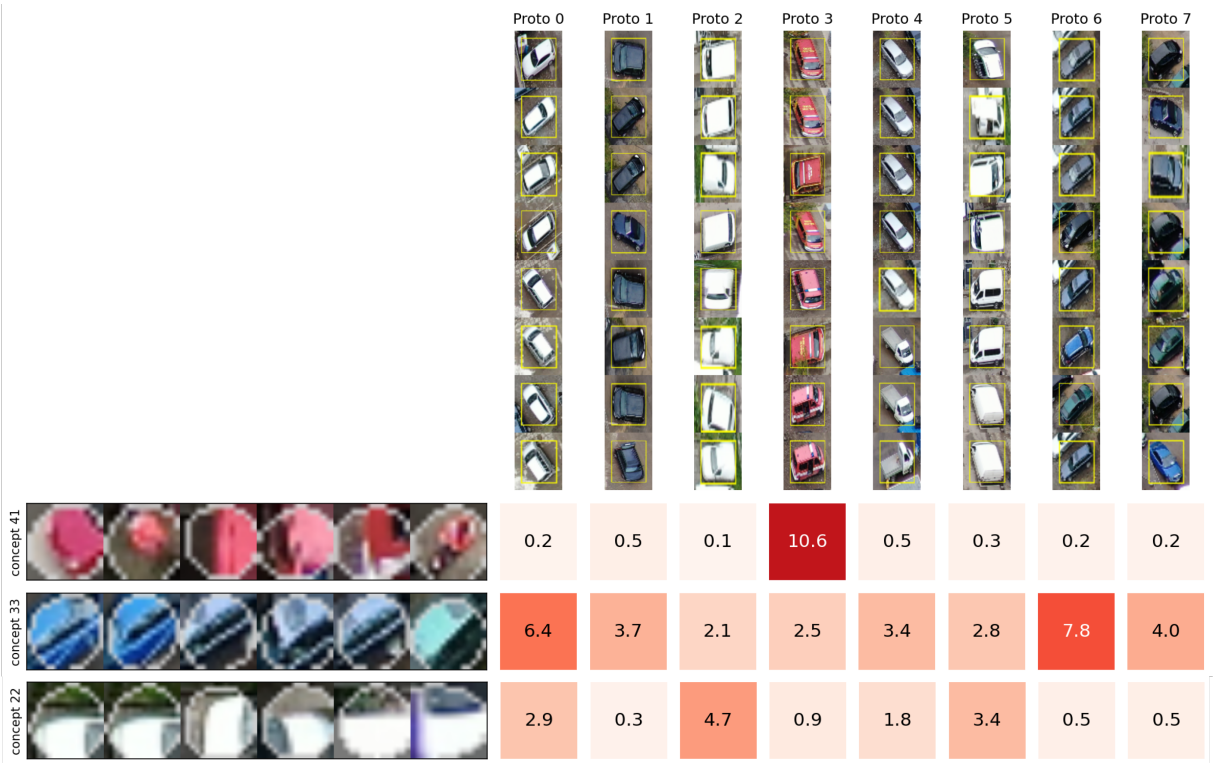


Figure 3. PCX prototypes visualization on the YOLOv6s6 car localization model from AUTH.

Additionally, we report in Table 1 the computation of time ratios for the local explanations (i.e., the raw attributions) w.r.t. the predictions, as well as the time ratios for the global explanations (i.e. the concept-based PCX explanations) w.r.t. all local explanations on the respective datasets for the AI models from AUTH. We can see that the KPIs match the TVs for 3 out of 4 ratios, thus fulfilling the main goals of the TEMA **Objective OA1 "Increase trustworthiness of extreme data analysis algorithms"**.

Table 1. Comparison of computation time ratios for local and global XAI methods between KPI and TV using the AI models from AUTH for segmentation and localization. These time ratios are reported as average over 100 samples and recorded on GPU.

Computation Time Ratio	Model	# Datapoints	KPI	TV
local XAI time / prediction time	U-Net	1055	2.99	4
	YOLOv6s6	695	5.97	4
global XAI time / all local XAI time	U-Net	1055	3.80	10
	YOLOv6s6	695	1.34	10

Overall, the explainability advances achieved by FHHI, both in terms of research output on generic XAI methods, as well as in key NDM use-cases, including fire classification, person and car detection, as well as segmentation of flood and burnt area, improve the transparency and robustness of AI models and support their reliable and near-real-time deployment in high-stake scenarios.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

3.3. XAI on diffusion models for image generation

SOTA (incl. TEMA M1-M18)

The explainability of diffusion models, particularly in image generation, is a rapidly evolving area in AI research with rapid advances, crucial for tasks like image annotation and automated labelling. These models synthesize images by iteratively de-noising random noise, a generative process that is inherently opaque and difficult to interpret. To address this, researchers have focused on visualizing the inner mechanisms of diffusion modelsspecifically, how visual features evolve and how attention is distributed throughout the generation process [61]. Tools that map attention across de-noising steps offer a clearer view of how image elements emerge, helping align model behaviour with human-understandable concepts.

Complementing visual techniques, LLMs are increasingly being used to explain diffusion outputs in natural language. Methods like X-IQE [62], for instance, generate textual descriptions that assess the realism, semantic alignment, and aesthetic quality of images created by text-to-image models. These human-readable evaluations make it easier to understand and validate the reasoning behind a models output, especially in scenarios where automatic labelling or human oversight is required.

Advances beyond SOTA: M19-M30

ATOS has been employing recent developments that go beyond earlier approaches such as Diffusion Attention Attribution Model (DAAM) [21] by introducing more sophisticated tools like Attention Map Diffusers [63]. These allow for detailed explainability in state-of-the-art diffusion models, including Stable Diffusion XL [64], Stable Diffusion 3.5 [65], and Flux.1 [66]. At ATOS, this technology is being leveraged to extract token-level attention maps during the generation process, enabling granular control and insight over image composition, object placement, and visual quality. For labelling applications, this is especially powerful: by mapping attention directly to prompt tokens, we can generate bounding boxes or highlight regions corresponding to specific concepts. For example, using Flux.1 with the prompt “A fire in a forest with rising smoke as seen from a drone,” attention maps for the tokens “fire” and “smoke” precisely localize those elements in the imageresulting in accurate, automated annotations that support scalable and interpretable image labelling pipelines (see Figures 4, 5, 6, 7, 8).

Decentralized Inference for Forest Fire Classification

SOTA (incl. TEMA M1-M18)

During natural disasters, centralized machine learning infrastructures may fail due to network outages or deliberate cyber-attacks. Collaborative distributed machine learning mitigates these risks by enabling training and inference across multiple nodes operating independently on the cloud. Current SOTA methods in collaborative distributed machine learning include federated learning techniques that facilitate decentralized training across multiple computational nodes using private datasets without raw data sharing [67, 68]. Additionally, teacher-student KD methods have been widely adopted for efficient knowledge transfer from large teacher models

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 4. Synthetic Image generated with Flux.1 of a forest fire with smoke.

Figure 5. Attention Map generated for the token "Fire" of the diffusion model

into compact student models, significantly improving computational efficiency and generalization performance [69, 70, 71, 72]. Advanced variations of KD, such as FitNets [73] and attention-based methods [74], have further refined the student model training process. Concurrently, OOD detection methods, such as those employing maximum softmax output probabilities [75], have been proposed to evaluate whether incoming data are significantly different from the model's training distribution, ensuring reliable inference in practical scenarios. However, despite these advancements, current frameworks typically do not integrate multiple distributed learning workflows such as federated learning, multi-teacher distillation, and distributed inference within a single, cohesive cloud-based infrastructure. This integration remains a crucial gap that limits

This project has received funding from the European Union – European Commission under grant agreement No 101093003

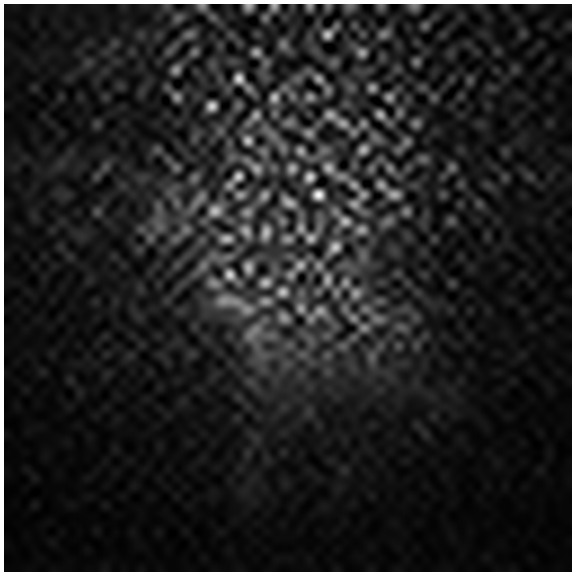


Figure 6. Attention Map generated for the token "Smoke" of the diffusion model

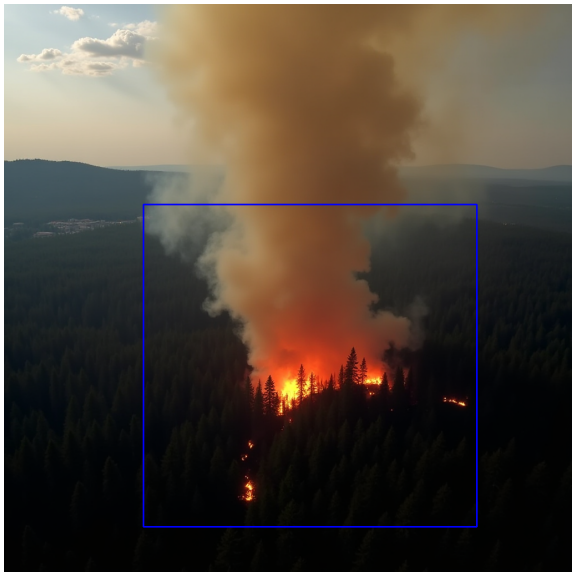


Figure 7. Resulting automatic bounding box for the class fire created from the attention map.

the versatility and robustness required in NDM.

As reported in D3.1, in order to address this critical gap, AUTH developed a decentralized DNN inference framework incorporating two novel methodologies: the Individualized Agent Model Selection Process (IAMSP) and the pioneering Quality of Inference (QoI) Byzantine Fault-Tolerant (BFT) consensus protocol. The IAMSP significantly improves recognition performance among decentralized agents by enabling collaboration and model optimization based on collective insights, exemplified by an accuracy increase of **16.69**% using EfficientNet B4 [76] on fire recognition tasks. The above-mentioned method proposed by AUTH has been published in the

This project has received funding from the European Union – European Commission under grant agreement No 101093003

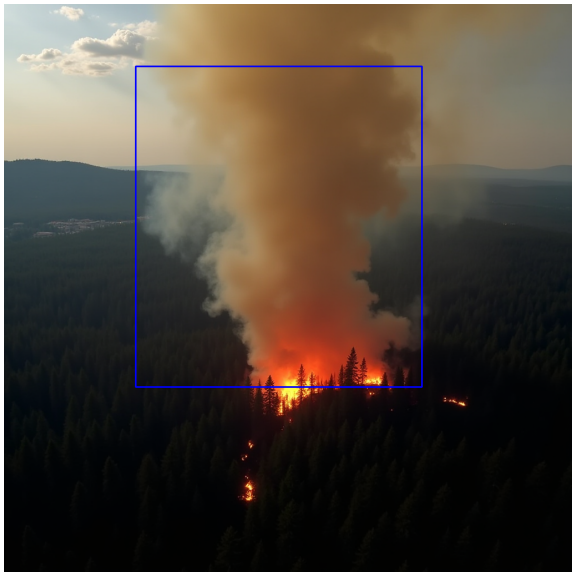


Figure 8. Resulting automatic bounding box for the class smoke created from the attention map.

"IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW)" [22].

Advances beyond SOTA

AUTH significantly advances beyond prior work by introducing a decentralized S-BFT consensus mechanism specifically tailored for D-DNN inference classification. Previously, AUTH developed the QoI protocol [22], a pioneering approach that addressed Byzantine fault tolerance but faced substantial scalability limitations as the number of participating DNN nodes increased.

To overcome these challenges, this work extends AUTHs earlier contributions by implementing a sharding-based consensus approach. This enhancement partitions the DNN network into independent sub-networks (shards), enabling decentralized inference at scale, thereby drastically reducing computational complexity and enhancing robustness against attacks, such as Distributed Denial of Service (DDoS) [77]. Furthermore, unlike previous methods, the new sharding mechanism incorporates OOD detectors to group nodes with similar domain expertise, significantly improving system accuracy and reliability. These novel contributions directly address the scalability and robustness gaps in previous AUTH solutions, setting a new benchmark for decentralized inference frameworks in critical scenarios, such as NDM. The developed method is described in detail in a conference paper [23]:

> D. Papaioannou, V. Mygdalis and I. Pitas, "*A Decentralized Sharding BFT Consensus Approach, for Efficient Decentralized DNN Inference Classification*", 2025 IEEE Symposium on Computers and Communications (ISCC) - 5th International Workshop on Distributed Intelligent Systems (DistInSys 2025), Bologna, Italy, 2025.

Figure 9 illustrates the developed decentralized S-BFT consensus architecture designed for D-DNN inference classification. The figure demonstrates how multiple decentralized DNN nodes are organized into distinct groups (shards). Each shard autonomously handles local consensus for classification inference tasks. An OOD detection mechanism ensures nodes within a shard

This project has received funding from the European Union – European Commission under grant agreement No 101093003

share similar domain expertise, optimizing the decision-making process and enhancing the systems robustness and scalability.

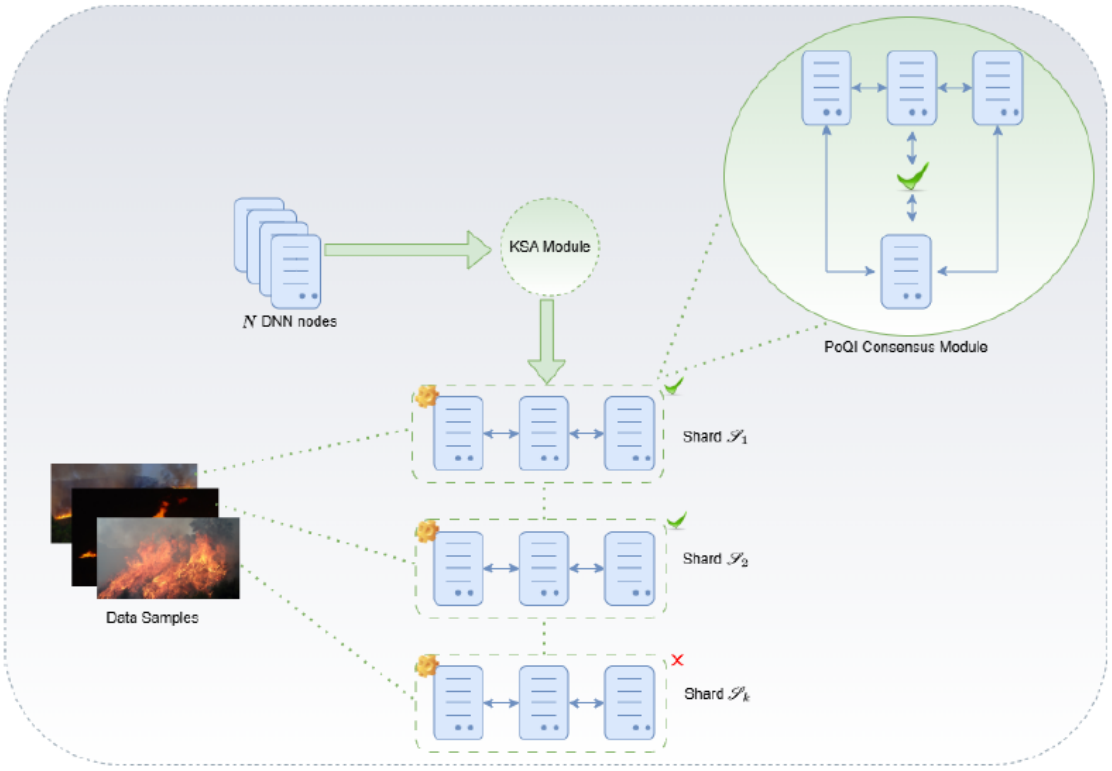


Figure 9. Decentralized S-BFT consensus architecture for D-DNN classification inference.

table 2 provides a comparative evaluation of the proposed S-BFT consensus protocol against standard aggregation strategiesAverage, Majority Voting, and Qolacross three benchmark datasets: CIFAR-10 [11], STL-10 [12], and the within TEMA-developed Blaze [13].

Table 2. Accuracy (%) comparison between the Shard-based Consensus Protocol and competing aggregation methods across Cifar10 [11], STL-10 [12] and Blaze [13] datasets, highlighting results obtained from one node.

Dataset	Average	Majority Voting	PoQI	SBFT (AUTH)
Cifar10 [11]	10.48	54.25	38.38	**92.24**
STL-10 [12]	11.15	50.64	41.77	**76.74**
Blaze [13]	9.97	57.52	43.46	**84.05**

The results clearly demonstrate the superiority of S-BFT in terms of classification accuracy, significantly outperforming competing methods in all evaluated scenarios. Notably, S-BFT achieves an accuracy of 92.24% on CIFAR-10, compared to only 54.25% for Majority Voting and 38.38% for QoI [22]. A similar trend is observed on STL-10 and Blaze, where the proposed S-BFT outperforms the closest alternative by margins exceeding 20%. This performance gain is attributed to S-BFT ability to isolate and exclude underperforming or irrelevant DNN nodes

This project has received funding from the European Union – European Commission under grant agreement No 101093003

through its Knowledge Self-Assessment (KSA) module, thereby forming task-relevant shards that foster consensus among only the most competent nodes. These findings validate the robustness and scalability of the S-BFT approach, especially under scenarios with potential Byzantine failures or non-Independent and Identically Distributed (IID) data distributions. This achievement surpasses the targeted "**Image recognition accuracy**" KPI of objective **OA2** "Increase accuracy of extreme data analysis algorithms".

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4. Real-time semantic visual analysis and remote sensing

4.1. Introduction

From M19 to M30 of the TEMA project, substantial progress was made in advancing AI-driven techniques for real-time semantic visual analysis and satellite-based remote sensing tailored to natural disaster management. D3.2 serves as an extension of D3.1, continuing the developments initiated in that phase. This work aligns with the objectives of WP3, specifically the ones associated with T3.2. It refines and scales up models introduced earlier, applies them to new disaster scenarios or introduces new methods that have not been tackled in D3.1.

A key focus was improving segmentation techniques for fire, smoke, and flood regions using both supervised and unsupervised approaches. Novel models integrating multi-modal data, such as RGB and infrared imagery, achieved robust real-time performance under complex conditions like sensor misalignment and occlusion. Unsupervised strategies were also developed to reduce dependence on labeled data in fast-evolving disaster contexts.

Complementary tools were introduced to support real-time situational awareness, including wildfire video summarisation, skeleton-based action recognition, and air quality forecasting. AUTH, DLR, and ATOS contributed to enhancing remote sensing pipelines for flood and fire monitoring. DLR's integration of direct broadcast from the Neustrelitz receiving station cut Sentinel-1 data latency by a factor of five and was validated during recent flood events in Europe.

Further advances included real-time object detection for aerial imagery, stereo-based road damage reconstruction, and fairness-enhanced detection of people and vehicles in flooded zones. Privacy-preserving modules for anonymising faces and license plates were also developed. Finally, 3D smoke dispersion modelling was improved using wind field estimation and finite element simulations, offering real-time insights to guide emergency response.

4.2. AI algorithms for visual data analysis

In Deliverable D3.2 (WP3, **T3.2**), we extend D3.1s methods by a suite of robust, real-time algorithms tailored to complex disaster scenarios involving wildfires and floods. These include innovations in supervised and unsupervised fire and smoke region segmentation, flood detection, privacy-preserving visual analysis, multi-agent learning for fire classification, and efficient 3D reconstruction for road damage assessment.

Additionally, AUTH introduced novel methodologies for wildfire video summarisation, skeleton-based action recognition, and air quality forecasting. Each solution is designed to meet the projects key objectives: improving algorithmic accuracy (**OA2**), enhancing responsiveness and computational efficiency (**OA3**), and enabling scalable, ethically responsible deployment in real-world NDM settings. The following subsections detail AUTHs advancements, evaluations, and alignment with the projects scientific and technical goals. The specific contributions are detailed in the following Sections.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4.2.1. Fire and smoke region segmentation

SOTA (incl. TEMA M1-M18)

During a wildfire, precise fire and smoke image region segmentation is essential for modelling fire spread and formulating effective firefighting strategies. During the first reporting period and as described in Deliverable D3.1, AUTH developed state-of-the-art fire segmentation methods by exploiting the RGB modality. During this period, AUTH focused on improving segmentation methods, by leveraging the IR modality, as well.

Fire and smoke can be imaged using a variety of light spectral bands, including visible, Short-Wave Infrared (SWIR), Mid-Wave Infrared (MWIR), and MWIR ones. There have been numerous methods proposed for fusing RGB and thermal modalities, with most applied to datasets like Pst900 [78] and MFNet [4]. These datasets usually contain well-registered RGB and thermal images with little occlusion between modalities. Currently, there are no multi-modal unregistered fusion strategies specifically designed for semantic region segmentation, apart from medical applications. Wildfire scenarios, where high-density smoke creates diverse and highly variable views between the two modalities, pose significant challenges for effective fusion.

Additionally, in the domain of fire and smoke segmentation, there are currently few efforts that simultaneously address both flame and smoke segmentation in a single model. Most existing works utilize relatively simple datasets, such as Corsican [14], which features clearly visible flame candidates, or unrealistic datasets like FLAME1 [79], with flames being represented in very unusual circumstances, such as within snow.

Advances beyond SOTA

AUTH significantly advances beyond existing SOTA methods in concurrent fire and smoke segmentation by introducing a novel intermediate image fusion architecture that effectively combines visible (RGB) and IR modalities. Unlike traditional single-modality techniques, AUTH developed a method that leverages attention mechanisms within a real-time DNN to dynamically fuse complementary features uniquely captured by each modality, addressing critical limitations such as dense smoke obscuration. The integration of a U-Net-inspired decoder enhances spatial reconstruction precision, providing more accurate segmentation outcomes even under challenging conditions like sensor misalignment and partial sensor failures. Notably, the lightweight DNN architecture delivers robust segmentation performance comparable to leading models on standard urban datasets and significantly surpasses them in wildfire-specific scenarios. These advances position the developed system as particularly suitable for real-time robotic deployment in dynamic, high-risk wildfire response operations. A technical report describing the details of the developed method has been submitted in the form of a journal paper [80]:

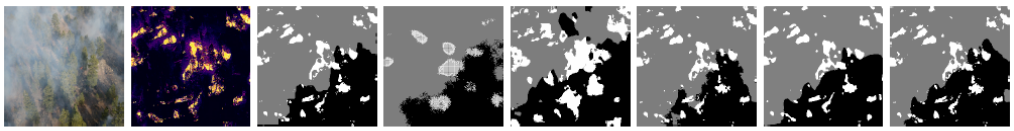
> D. Fotiou, V. Mygdalis and I. Pitas, "RoboFireFuseNet: Robust Fusion of Visible and Infrared Wildfire Imaging for Real-Time Flame and Smoke Segmentation", technical report, 2025,

Figure 10 depicts a qualitative comparison of the developed method with several SOTA fusion-based fire and smoke region segmentation architectures.

Table 3 demonstrates the comparison of several SOTA region segmentation methods on BG, fire, and smoke classes in terms of Recall and IoU. The proposed AUTH method achieves superior performance across most metrics.

From both Figure 10 and Table 3, it can be derived that the proposed method outperforms other

This project has received funding from the European Union – European Commission under grant agreement No 101093003



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 10. Qualitative evaluation on the FLAME2 dataset [1]: (a) RGB input, (b) IR input, (c) Ground Truth, (d) RTFNet [2], (e) EGFNet [3], (f) MFNet [4], and (g) Sigma-T [5] (h) AUTH proposed method. The gray color represents the smoke class, while the white color represents the fire class.

Table 3. Comparison of region segmentation methods on BG, fire, and smoke classes in terms of Recall and IoU. The developed RFFNet achieves superior performance across most metrics.

Method	BG Recall (%)	BG IoU (%)	Fire Recall (%)	Fire IoU (%)	Smoke Recall (%)	Smoke IoU (%)	Avg Recall (%)	mIoU (%)
PIDNet-RGB [6]	93.51	90.96	52.13	19.54	81.35	71.14	75.66	61.21
PIDNet-IR [6]	86.76	83.68	89.17	47.66	73.21	44.81	83.05	58.71
PIDNet-Early [6]	95.46	93.65	81.59	52.34	84.64	74.1	88.25	73.90
MFNet [4]	95.45	92.68	96.08	65.70	89.06	82.40	93.53	80.26
RTFNet [2]	95.12	76.77	37.36	30.36	89.13	76.25	73.87	65.42
GMNet [81]	88.63	83.65	15.39	15.39	72.04	63.20	67.53	54.08
EGFNet [3]	91.29	88.73	49.52	21.05	81.93	73.16	74.27	60.98
Sigma-T [5]	96.67	93.93	92.54	78.727	88.38	86.12	92.96	86.27
RFFNet - AUTH method	**98.2**	**95.08**	**93.85**	**81.5**	**91.08**	**87.98**	**94.37**	**88.17**

SOTA fusion-based fire and smoke region segmentation strategies in all metrics and classes. The main difference is observed in fire class. Fire is difficult to be segmented, because it is mostly under dense smoke areas and is contained of small and sparse fire regions. AUTH's model with the exploitation of skip connections gives baseline architecture significant boost on recall and mIoU of that class. The proposed method satisfies the objective **OA2** "Increase accuracy of extreme data analysis algorithms" as it consistently outperforms several previous SOTA fire and smoke region segmentation models. It is also compliant with the objective **OA3** "Increase Responsiveness/Speed of Extreme Data Analysis Algorithms" by delivering rapid and accurate fire and smoke region segmentation in real time, meeting the "**Visual analysis speed**" KPI.

4.2.2. Unsupervised fire region segmentation

SOTA (incl. TEMA M1-M18)

Unsupervised fire region segmentation is particularly valuable in NDM due to the prevalent lack of annotated datasets specifically tailored for fire events. Traditional supervised approaches require extensive manual labeling, which is time-consuming and impractical during rapidly evolving disasters. By leveraging unsupervised segmentation, regions affected by fire can be rapidly and accurately identified without dependence on pre-existing labeled data.

Unsupervised region segmentation methods have gained substantial attention due to their ability to reduce reliance on extensive pixel-level annotations, leveraging strategies such as mutual information maximisation [82], invariance and equivariance constraints [83], and contrastive learning [84, 85, 86]. Particularly, recent approaches such as Self-supervised Transformer with Energy-based Graph Optimization (STEGO) [84] have significantly advanced USS performance by distilling self-supervised self-DIstillation with NO labels (DINO) visual features [24] into coherent semantic clusters. Despite these developments, unsupervised region segmentation methods still face notable limitations, including the improper initialisation of class prototypes, the need for manual or supervised hyperparameter tuning, and poor performance in scenarios involving smaller or less prevalent objects. Additionally, these methods often require supervised

This project has received funding from the European Union – European Commission under grant agreement No 101093003

post-processing to map clusters to meaningful class labels, limiting their practical utility and deployment efficiency.

Advances beyond SOTA

AUTH contributes to addressing the unsupervised fire region segmentation challenges by introducing a novel method leveraging minimal pixel-based prompting and dynamic thresholding mechanisms, called PXL-USS. A manuscript with the details of the proposed method has been submitted in the form of a journal paper [87]:

> M. Tzimas, V. Mygdalis, C. Papaioannidis, I. Pitas, "Extreme Weakly Supervised Binary Semantic Image Segmentation via One-Pixel Supervision", technical report, 2025.

The developed PXL-USS method significantly alleviates the challenges posed by the absence of extensive annotated datasets for fire events. By using sparse pixel annotations as class prototypes, AUTH ensures precise initialisation and effective clustering of fire-related features within the self-supervised DINO [24] feature space. Moreover, the dynamic computation of contrastive loss thresholds eliminates the need for manual hyperparameter tuning, enhancing model robustness and deployment efficiency. Consequently, AUTH approach bridges the performance gap between unsupervised and supervised segmentation methods, enabling rapid, accurate, and scalable segmentation of fire regions, critical for effective NDM. Figure 11 depicts fire region segmentation masks generated by the proposed PXL-USS method. It is obvious that PXL-USS (bottom) closely resembles ground truth masks (middle) in terms of visual quality.

Figure 11. Fire region segmentation masks (bottom) generated by the proposed PXL-USS method. The results demonstrate that PXL-USS closely resembles ground truth masks (middle) in terms of visual quality.

Table 4 demonstrates the mIoU results on the Corsican Fire Database [14], with models arranged in ascending order of their data requirements.

Remarkably, AUTH approach achieves a **+17**% improvement in mIoU compared to STEGO-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 4. mIoU comparison of various unsupervised segmentation models on the Corsican Fire Database [14]. The proposed PXL-USS consistently demonstrates superior performance compared to previous SOTA unsupervised methods.

Model	mIoU
STEGO Unsupervised [84]	60.75
DINO [24] + Prototypes (4 Imgs * 4 Points)	56.87 ± 0.49
DINO [24] + Prototypes (BEST)	58.3 ± 1.12
AUTH method	
PXL-USS (4 Imgs * 4 Points)	77.51 ± 0.87
PXL-USS (BEST)	78.23 ± 0.59

Unsupervised [84], despite being run only once, whereas STEGO required multiple training sessions to optimize its hyperparameters. Thus, it greatly surpasses the target value of 5% for the "**Semantic segmentation accuracy**" KPI of objective **OA2** "Increase accuracy of extreme data analysis algorithms".

4.2.3. Wildfire video summarisation

SOTA (incl. TEMA M1-M18)

Video summarisation is particularly useful for NDM because it enables rapid review and analysis of vast amounts of video footage, a critical capability during disaster events such as wildfires. Given the exponential growth of available video data, authorities managing natural disasters must quickly extract essential information, identify urgent situations, and make informed decisions. Early video summarisation techniques predominantly employed clustering or dictionary learning methodologies [88]. Despite their initial effectiveness, these methods struggled to capture high-level semantic content, such as narrative consistency. Subsequent deep learning approaches adopted supervised Long Short-Term Memory (LSTM) networks [89, 90, 91] and unsupervised adversarial models [92, 93, 94] to better model inter-frame dependencies. Transformers were later introduced to directly model long-range dependencies in videos, with methods such as VASNet [95] and DSNet [96] significantly advancing the field. However, existing transformer-based methods implicitly assume the necessity of capturing long-range dependencies, potentially resulting in redundant computations and heavy computational demands [97]. This assumption remains largely unverified, highlighting a critical gap: the need for efficient architectures capable of explicitly modelling short-range dependencies without sacrificing summarisation quality or computational efficiency.

Advances beyond SOTA

AUTH developed the DIVide and SUMmarize (DIV-SUM) framework introduces two significant advances over current SOTA video summarisation methods. The developed DIV-SUM method is described in a conference paper [98]:

> E. Charalampakis, C. Papaioannidis and I. Pitas, "Divide-and-Summarize: Enhancing Deep Neural Video Summarization", International Conference on Image Processing (ICIP), 2025.

DIV-SUM implements a novel video fragmentation mechanism, partitioning input videos into smaller, fixed-length segments. This approach explicitly models short-range inter-frame dependencies, leading to improved summarisation performance and significantly reduced inference

This project has received funding from the European Union – European Commission under grant agreement No 101093003

costs, as fragments can be processed in parallel. Secondly, the introduction of a target space quantisation module converts the regression task of predicting continuous frame importance scores into a simpler classification problem, thereby enhancing generalisation and robustness. Extensive experimental evaluation demonstrates that DIV-SUM achieves SOTA performance on the SumMe benchmark [15] and remains competitive on TVSum [16], all while significantly decreasing computational overhead compared to global-field-of-view Transformer-based methods such as PGL-SUM [97] and DSNet [96].

Figure 12 illustrates the inference pipeline of the DIV-SUM method. It shows how the input video depicting fire sources and dense smoke is first fragmented into smaller segments, which are independently processed by separate summarisation modules in parallel. Each summarizer module generates frame-level importance predictions for its respective segment. The outputs from all modules are then concatenated to produce the final summarized representation of the entire video, significantly reducing computational overhead compared to traditional methods that process videos in their entirety.

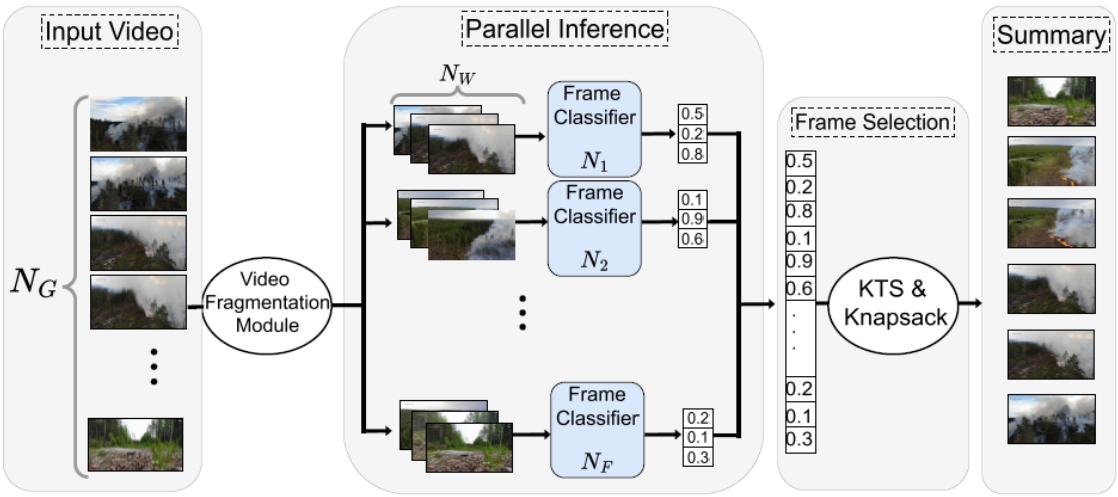


Figure 12. DIV-SUM inference pipeline.

Table 5 presents a performance comparison (measured by average F1-Score percentages) of DIV-SUM against various recent SOTA supervised video summarisation methods, including DR-DSN [96], SUM-FCN [90], VASNet [95], and others, on the canonical datasets SumMe [15] and TVSum [16].

DIV-SUM achieves the highest accuracy on the SumMe dataset (60.51%), competitive performance on TVSum (60.53%), and the best overall average accuracy (60.52%), demonstrating its robustness and effectiveness across datasets compared to prior approaches.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 5. Comparison of supervised video summarisation performance on SumMe [15] and TVSum [16] datasets. Higher values indicate better performance.

Method	SumMe	TVSum	Average
DR-DSN [96]	42.1	58.1	50.1
SUM-FCN [90]	47.5	56.8	52.15
VASNet [95]	49.7	61.4	55.55
A-AVS [99]	43.9	59.4	51.65
M-AVS [99]	44.4	61.0	52.7
LMHA [100]	51.1	61.0	56.05
CAAN [101]	50.6	59.3	54.95
VSS-Net [102]	51.5	61.0	56.25
DSNet [96]	51.2	61.9	56.55
PGL-SUM [97]	55.6	61.0	58.3
DIV-SUM (proposed)	**60.51**	60.53	**60.52**

4.2.4. Air quality timeseries forecasting

SOTA (incl. TEMA M1-M18)

Air quality timeseries forecasting is crucial for NDM in wildfire scenarios, as it enables early prediction and mitigation of fire-related risks. By analysing historical air quality dataincluding pollutants like nonmethane hydrocarbons ($NMHC$), benzene (C_6H_6), carbon monoxide (CO), and nitric oxides (NO_x, NO_2), DNN models can forecast hazardous air conditions linked to wildfires. This information aids authorities in anticipating fire spread, assessing air quality impact, and planning evacuations. Additionally, accurate forecasting enhances resource allocation for firefighting efforts and enables proactive health advisories for affected communities, improving overall disaster response efficiency. Existing time series forecasting methods rely on a variety of deep learning approaches, including Recurrent Neural Networks (RNNs) [103], LSTM networks [104], sequence-to-sequence architectures [105, 106], and attention-based models [107, 108]. While these models have achieved strong performance in general forecasting tasks, they often struggle in causal forecasting scenarios, where an arbitrary subset of variables serves as the forecasting target while the remaining variables act as exogenous inputs. Traditional methods such as AutoRegressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA) models [109, 110] have been surpassed by DNNs in evaluation metrics, but these DNNs still face challenges in capturing interdependencies between input channels. Attempts to improve causal forecasting include spatial-temporal attention mechanisms [111], Graph Neural Networks (GNNs) [112], and hybrid approaches like LSTNet [113], but they often fail to generalize well across datasets and suffer from limitations in multistep forecasting [114].

Advances beyond SOTA

To address the limitations of previous SOTA methods, AUTH developed the Generative-Regressing Recurrent Neural Network (GRRNN), a novel framework that synergistically integrates generative representation learning and regression to enhance causal time series forecasting. The developed method has been published in the "IEEE Transactions on Artificial Intelligence" journal [115]:

This project has received funding from the European Union – European Commission under grant agreement No 101093003

- G. Chatziparaskevas, I. Mademlis, and I. Pitas, "Generative representation learning in recurrent neural networks for causal timeseries forecasting", IEEE Transactions on Artificial Intelligence, vol. 5, pp. 64126425, 2024.

The key contributions of GRRNN are: (1) a Wasserstein recurrent conditional Generative Adversarial Network (WRCGAN) that transforms historical exogenous data into a more meaningful latent representation, improving feature extraction; (2) a sequence-to-sequence regression module that utilizes these rich features for accurate long-term forecasting; and (3) a joint end-to-end multitask training strategy, which leverages adversarial learning to refine the learned representations and improve forecasting robustness. Experimental results demonstrate that GRRNN significantly outperforms SOTA models in multistep causal forecasting, validating its effectiveness in real-world applications.

Table 6 demonstrates the forecasting error rates for the test set of the Air Quality dataset [17]. The Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (SMAPE) and the Root Mean Squared Error (RMSE) metrics are used. A seven timesteps (long/multistep) forecasting horizon is used.

Table 6. Comparison of forecasting error rates for the Air Quality dataset [17]. Lower values indicate better performance.

Method	MAE	SMAPE	RMSE
Enc–Dec [113]	0.161	0.587	0.209
DARNN [111]	0.147	0.622	0.205
DSTP-RNN [116]	0.158	0.746	0.224
Stem-GNN [117]	0.171	0.603	0.211
GRRNN	**0.133**	**0.471**	0.174
GRRNN-T	**0.133**	0.491	**0.173**

Overall, the developed GRRNN architecture is typically either the best one or the second best one across almost all evaluation metrics. It is intended as a multiple-step-ahead forecasting method and, indeed, performs best in terms of error rates in the challenging experimental setting that has a long forecast horizon of seven timesteps.

4.2.5. Flood region segmentation

SOTA (incl. TEMA M1-M18)

Accurate flood region segmentation plays a critical role in NDM, facilitating timely monitoring and precise decision-making to support effective response strategies and mitigate widespread impacts. SOTA DNN architectures for semantic region segmentation have been utilized to extract water regions from surveillance images, notably for estimating river water levels [118]. Systems like V-FloodNet employ video segmentation techniques, being able to provide real-time flood monitoring [119]. H2O-Net [120], leveraging self-supervised learning and adversarial domain adaptation, eliminates the need for hand annotations while excelling in segmenting high-resolution satellite imagery. Additionally, UAV-based segmentation methods optimized for edge computing enable autonomous flood mapping in disaster scenarios [121].

As reported in D3.1, AUTH introduced a novel flood region segmentation dataset called FloodSeg, containing annotated images sourced from diverse datasets. It serves as a benchmark for evaluating SOTA flood region segmentation DNN architectures. Leveraging the ST++

This project has received funding from the European Union – European Commission under grant agreement No 101093003

self-training method [122] with the PSPnet architecture [123], the novel AUTH flood region segmentation methods demonstrated promising improvements towards reaching the target value of 5% for the "**Semantic segmentation accuracy**" KPI of objective **OA2** "Increase accuracy of extreme data analysis algorithms", achieving a **3.5%** mIoU increase over previous benchmarks. Additionally, AUTH solution processes images at a speed of **79** Frames Per Second (FPS) on a GeForce GTX 1080 GPU, which is much faster than "**Visual analysis speed**" KPI of objective **OA3** "Increase responsiveness/speed of extreme data analysis algorithms". This work is described in a conference paper [25]:

> A. Gerontopoulos, D. Papaioannou, C. Papaioannidis and I. Pitas, "Real-Time Flood Water Segmentation with Deep Neural Networks", 2025 IEEE 25th International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW), Bologna, Italy, 2025.

Existing flood region segmentation approaches often overlook a fundamental challenge of the flood region segmentation problem. That is, there is limited variability in the foreground (flooded) region and substantial variability in the background elements, resulting in a broader feature distribution that can degrade segmentation performance.

Advances beyond SOTA

To deal with the imbalanced variance in visual appearance between the foreground (flooded regions) and the background, AUTH developed a novel method employing Self-KD, where a Teacher model is trained on augmented inputs specifically designed to reduce background variance and align it more closely with the foreground. The developed method is described in detail in a conference paper [26]:

> P. Mentesidis, V. Mygdalis and I. Pitas, "Improve Real-Time Flood Segmentation by Encoding and Distilling Foreground Information", International Conference on Image Processing (ICIP), 2025,

During Student training, the Teacher processes these augmented inputs while the Student learns directly from real-world data, enabling the Student to focus on discriminative features relevant to flood regions. This targeted learning strategy effectively suppresses the influence of noisy background areas, enhancing the Student models ability to distinguish reliably between flooded and non-flooded regions, even in the presence of significant clutter and variability. The proposed method demonstrates consistent performance improvements across SOTA real-time semantic segmentation models, highlighting its effectiveness and robustness. It is model-agnostic and operates exclusively during the training phase, ensuring no additional computational overhead during inference. Consequently, it is particularly suitable for real-time applications where speed and efficiency are paramount. This advancement not only enhances the effectiveness of real-time flood segmentation models but also demonstrates the flexibility of the proposed methodology, establishing it as a valuable tool for training robust systems in natural disaster management. Figure 13 displays examples of flood region segmentation on FloodSeg images [25] using the PIDNet [6] trained with the Self-KD framework developed by AUTH.

Table 7 demonstrates mIoU results for SOTA flood region segmentation architectures on the FloodSeg dataset. It is clear that the proposed Self-KD framework consistently improves segmentation performance compared to training with Feature-based KD (Feature-KD) or without any Self-KD (Base).

The developed method applied on the PIDNet [6], achieves an **1%** mIoU increase over the

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 13. Flood region segmentation masks (in purple) overlaid on FloodSeg images. The masks were generated by the PIDNet [6] trained with the novel Self-KD framework proposed by AUTH.

Table 7. mIoU results for SOTA flood region segmentation architectures on the FloodSeg dataset. The proposed KD framework consistently improves segmentation performance compared to training with Feature-based KD (Feature-KD) or without any Self-KD (Base).

Model	mIoU			Comparison	
	Base	Feature-KD	Proposed	vs Base	vs Feature-KD
BiseNetV2 [124]	0.832	0.824	**0.835**	+0.003 (+0.3%)	+0.003 (+0.3%)
PPLiteSeg [125]	0.822	0.825	**0.841**	+0.019 (+2.3%)	+0.016 (+1.9%)
DDRNet [126]	0.825	0.819	**0.828**	+0.003 (+0.3%)	+0.009 (+1.0%)
PIDNet [6]	0.853	0.868	**0.875**	+0.022 (+2.5%)	+0.007 (+0.8%)

previous AUTH method described in D3.1 on the FloodSeg dataset, leading to a total **4.5%** mIoU increase over previous methods, satisfying the objective **OA2** "Increase accuracy of extreme data analysis algorithms". AUTH solution processes images in real-time satisfying the "**Visual analysis speed**" KPI of objective **OA3** "Increase responsiveness/speed of extreme data analysis algorithms".

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4.2.6. Road surface 3D reconstruction for damage assessment

SOTA (incl. TEMA M1-M18)

Stereo matching for road surface 3D reconstruction is highly relevant for NDM, as it enables accurate assessment of road damage, terrain deformation, and infrastructure stability after disasters such as floods. Previous SOTA methods for stereo matching in road scenarios often struggle with generalisation to unseen environments, particularly in challenging conditions such as varying illumination, occlusions, and dynamic obstacles [127, 128, 129]. While traditional approaches rely on handcrafted feature extraction and cost aggregation techniques [130, 131], they lack the robustness needed for complex real-world road scenes. Recent deep learning-based stereo matching models have improved disparity estimation accuracy [132, 133, 134], but many suffer from overfitting to specific datasets and require large-scale annotated training data, which is often unavailable for disaster-affected environments [135, 136]. Additionally, few existing methods explicitly address the domain gap between training and real-world deployment, limiting their adaptability in dynamic or disaster-stricken road conditions [137, 138].

Advances beyond SOTA

To address the generalisation limitations of previous methods, AUTH introduces Decisive Disparity Diffusion Stereo (D3Stereo). The details of the proposed method have been published in the IEEE Transactions on Image Processing journal [7]:

> C. Liu, Y. Zhang, Q. Chen, I. Pitas, and R. Fan, "These maps are made by propagation: Adapting deep stereo networks to road scenarios with decisive disparity diffusion", IEEE Transactions on Image Processing, vol. 34, pp. 15161528, 2025.

D3Stereo is a decisive disparity diffusion strategy, which significantly advances stereo matching for road scenarios by addressing the limitations of existing deep learning-based methods. Unlike prior approaches that struggle with adaptive cost aggregation and disparity refinement, D3Stereo incorporates: (1) a recursive bilateral filtering algorithm for efficient and adaptive cost aggregation, (2) an intra-scale disparity diffusion algorithm for sparse disparity map completion, and (3) an inter-scale disparity inheritance algorithm for fine-grained disparity estimation at higher resolutions. Additionally, AUTH introduces a new dataset designed to comprehensively evaluate stereo matching-based road surface 3D reconstruction methods, addressing the lack of benchmark diversity in previous studies. Experimental results demonstrate that D3Stereo effectively adapts pre-trained deep learning models for stereo matching in both road and general scenes, outperforming existing methods in robustness, accuracy, and generalisation.

The quantitative and qualitative experimental results on the proposed UDTIRI-Stereo dataset are presented in Table 8 and Figure 14, respectively. UDTIRI-Stereo dataset [7] consists of 3,000 pairs of stereo images (resolution: $720 \times 1,280$ pixels), along with their disparity ground truth, collected across 12 scenarios under different illumination conditions (middle sunlight, intense sunlight, and street lighting at dark), weather conditions (tidy and watered), and road materials (asphalt and cement). Random 2D Perlin noise and digital twins of real-world pothole are introduced to the road data.

It is noticeable that when applying the proposed D3Stereo strategy to the existing stereo matching algorithms, End Point Error (EPE) and Percentage of Error Pixels (PEP) decrease by up to 63.10% and 83.26%, respectively. Although CreStereo demonstrates comparable performance

This project has received funding from the European Union – European Commission under grant agreement No 101093003

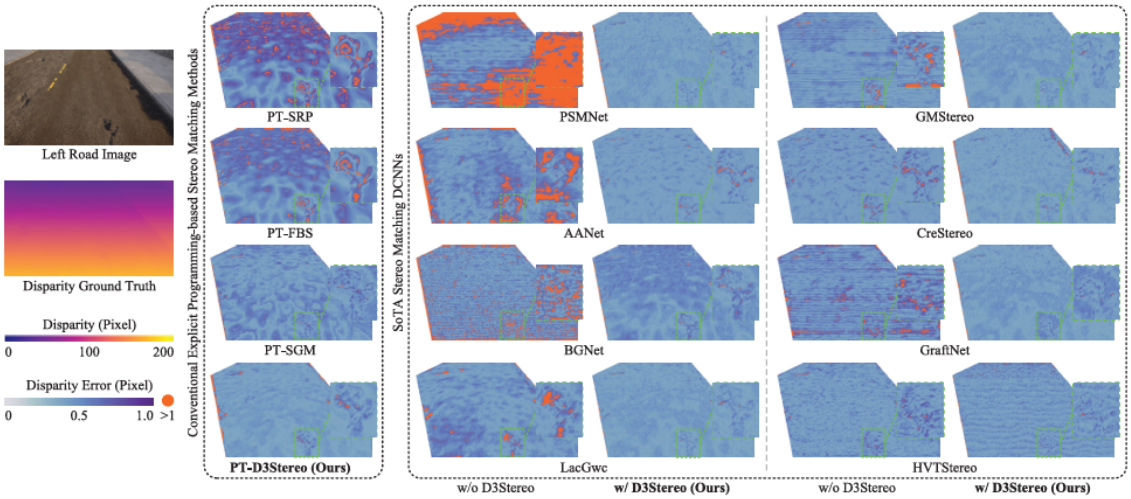


Figure 14. Examples of disparity estimation results the created UDTIRI-Stereo dataset [7].

in PEP with $\delta = 1$ when adapted to the UDTIRI-Stereo dataset using D3Stereo strategy, it shows dramatic improvement in PEP with $\delta = 0.5$ and EPE. Moreover, PT-D3Stereo diffuses decisive disparities across the entire image in a multi-directional fashion. This results in significantly improved disparity estimation results and a more uniform distribution of errors compared to other explicit programming-based stereo matching algorithms.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

*Table 8. Experimental results on the UDTIRI-Stereo dataset [7]. The best results are shown in **bold**.*

Method	PEP (%)	EPE (pixel)	PSNR (dB)	MSE	SSIM
(a) Explicit programming-based disparity estimation algorithms					
PT-SRP [139]	43.2	1.27	33.01	52.12	0.867
PT-FBS [140]	33.7	0.75	32.87	48.80	0.895
PT-SGM [136]	7.17	0.72	32.50	**41.12**	0.932
PT-D3Stereo (AUTH)	**5.90**	**0.65**	**33.14**	42.58	**0.951**
(b) Comparisons of SOTA stereo matching networks with and without D3Stereo					
Method	**PEP (%)**		**EPE (pixel)**	**PSNR (dB)**	**MSE**	**SSIM**
	$\delta=0.5$	$\delta=1$				
PSMNet [134]	46.7	13.2	0.84	32.36	55.51	0.894
PSMNet+D3Stereo (AUTH)	**4.81**	**2.21**	**0.31**	**34.75**	**32.80**	**0.953**
AANet [141]	35.4	8.69	0.56	34.01	38.14	0.932
AANet+D3Stereo (AUTH)	**11.1**	**1.79**	**0.43**	**34.21**	**36.49**	**0.948**
BGNet [142]	12.7	1.51	0.26	34.59	36.10	0.948
BGNet+D3Stereo (AUTH)	**3.53**	**1.20**	**0.21**	**34.72**	**34.07**	**0.954**
LacGwc [143]	18.6	2.87	**0.36**	34.45	36.91	0.945
LacGwc+D3Stereo (AUTH)	**4.61**	**2.01**	0.37	**34.61**	**34.28**	**0.953**
GMStereo [144]	23.6	4.12	0.43	32.66	47.39	0.937
GMStereo+D3Stereo (AUTH)	**3.91**	**1.46**	**0.31**	**34.66**	**33.48**	**0.953**
CreStereo [145]	6.02	1.40	0.34	34.77	34.85	0.950
CreStereo+D3Stereo (AUTH)	**4.58**	**1.10**	**0.30**	**34.76**	**33.87**	**0.952**
GraftNet [146]	20.9	2.78	0.33	34.72	36.69	0.943
GraftNet+D3Stereo (AUTH)	**4.73**	**1.57**	**0.32**	**34.57**	**32.35**	**0.951**
HVT Stereo [147]	6.83	1.95	0.36	34.79	31.11	0.951
HVT Stereo+D3Stereo (AUTH)	**4.88**	**1.30**	**0.34**	**34.79**	**31.11**	**0.951**

4.2.7. Person and car detection in flooded areas

SOTA (incl. TEMA M1-M18)

As reported in deliverable D3.1, AUTH has utilized and optimized YOLOv6 [148] models specifically for the critical task of identifying and localising humans and vehicles in flooded regions. These models are designed to efficiently detect and localize objects in high-resolution images, ensuring accurate identification of humans being in danger. AUTH YOLOv6 object detection models were trained in novel flood image datasets, processing images at **33** FPS on a GeForce GTX 1080 GPU satisfying the "**Visual analysis speed**" KPI of objective **OA3** "Increase responsiveness/speed of extreme data analysis algorithms".

Objects such as people and cars appear in diverse shapes and sizes due to variations in drone flight heights, viewing angles, and perspectives. Additionally, the shape of a person can vary significantly depending on different actions or postures. This variability poses significant challenges for person and car detection models as the traditional $L1$ loss used for training such models disproportionately penalizes predictions involving larger bounding boxes, leading to training instabilities.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Advances beyond SOTA

To mitigate training instabilities accompanying standard person and car detection models, AUTH developed a coordinate-based weighting strategy for the $L1$ loss, assigning higher error weights to smaller bounding boxes, balancing their contribution during training. The details of the developed method are described in detail in a conference paper that has been accepted for publication [27]:

> M. Tzimas, V. Mygdalis and I. Pitas, "A Weighting Loss Approach for Transformer-Based Object Detection", International Joint Conference on Neural Networks, Rome, Italy, 2025.

By integrating this approach into DEtection TRansformer (DETR)-based models such as RT-DETR [8], the method stabilizes optimisation, improves alignment between localisation losses, and enhances the overall detection accuracy for complex scenarios involving significant bounding box size disparities such as those encountered in person and car detection tasks.

Figure 15 displays the detection of persons in red and cars in green, in the flooded German Ahr valley, using RT-DETR-R18 trained with the proposed LSB method, which gives extra weight to tiny boxes during training, so the network learns not to ignore small, distant people and cars.

Figure 15. Person (red) and car (green) detection in the flooded Ahr valley, using RT-DETR-R18 [8] trained with the proposed method (LSB).

Table 9 demonstrates a mAP comparison of various real-time object detection models evaluated on the VisDrone dataset [18]. This dataset includes 10 object categories, such as pedestrians, cars, and bicycles, with images captured by drones.

Due to VisDrone [18] diverse backgrounds and the presence of small objects, VisDrone [18] provides a challenging benchmark for assessing the performance of object detection methods. It consists of 6,471 images for training, 548 images for validation, and 1,611 images for testing. It is obvious that RT-DETR-R18 [8] trained with the proposed method (LSB) consistently outperforms several other You Only Look Once (YOLO)-based models in terms of both mAP and mAP50 metrics. The proposed RT-DETR-R18+LSB satisfies the objective **OA2** "Increase accuracy of extreme data analysis algorithms" as it consistently outperforms several previous SOTA YOLO-based models. It is also compliant to the objective **OA3** "Increase Responsiveness/Speed of Extreme Data Analysis Algorithms" by delivering rapid and accurate person/car detection in flooded regions at real time, meeting "**Visual analysis speed**" KPI.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 9. mAP comparison of various real-time object detection models evaluated on the VisDrone dataset [18]. RT-DETR-R18 [8] with the proposed method (LSB) consistently outperforms several other YOLO-based models in terms of both mAP and mAP50 metrics.

Model	mAP	mAP50	Params (M)
YOLOv4-S [149]	14.6	26.2	9.1
YOLOv8-nano [150]	14.8	26.7	3.1
YOLOX-S [151]	14.3	26.2	8.9
YOLOv5-S [152]	17.2	30.8	9.1
YOLOv7-tiny [153]	15.0	27.4	6.1
YOLOv8-S [150]	18.1	31.6	11.1
YOLOv5-M [152]	19.6	33.9	25
RT-DETR-R18 [8]	20.4	35.6	19.8
RT-DETR-R18-*LSB*	**22.2**	**40.1**	**19.8**

4.2.8. Visual privacy preservation

SOTA (incl. TEMA M1-M18)

Ethical and privacy challenges have become a central concern in AI applications. This is especially relevant for tasks involving visual data, where individuals may be identifiable. A prominent example is facial recognition technology, which has raised significant privacy concerns. To mitigate such risks, various methods for face blurring and de-identification have been introduced [154], aiming to obscure personal identity in visual datasets while preserving the utility of the data.

Within M1-M18 of TEMA, AUTH introduced a a novel technique called Privacy via Adversarial Reprogramming (PAR). It leverages adversarial reprogramming to obfuscate sensitive information, such as facial identity, by modifying input images in a way that makes them unintelligible to both humans and standard facial recognition systems. Despite the applied perturbations, a target DNN remains capable of interpreting the image correctly.

Advances beyond SOTA

More recently, AUTH developed a novel privacy-preserving method based on the CenterFace [9] model to address concerns around facial data exposure in surveillance and emergency response scenarios. Leveraging CenterFaces lightweight and efficient face detection and alignment capabilities, the method detects and accurately localizes facial regions in real time, even on low-power edge devices. Once detected, facial areas can be selectively blurred or anonymized, ensuring individuals identities remain protected while still enabling situational awareness during natural disasters. This approach provides a practical balance between operational effectiveness and privacy, making it suitable for deployment in public monitoring systems, shelters, and disaster response efforts. This method is also fully compliant with objective **OA3** "Increase Responsiveness/Speed of Extreme Data Analysis Algorithms," as it enables rapid and efficient face detection and anonymisation, even in high-throughput, real-time disaster monitoring scenarios. Figure 16 depicts an example of visual privacy preservation using CenterFace [9].

In addition, AUTH developed a robust, privacy-preserving license plate detection system based on the YOLOv11-small architecture [155]. The team curated and preprocessed a large-scale dataset of over 350,000 annotated vehicle images, applying techniques such as duplicate re-

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 16. Visual privacy preservation using CenterFace [9].

moval, label correction, format unification, and incremental difficulty scoring. To improve model generalisation and reduce false positives, background images were incorporated into the training pipeline, and incremental learning was applied across Easy, Medium, and Hard subsets. A key innovation is the implementation of efficient anonymisation modules that automatically detect and blur license plates in both images and videos, supporting batch processing and achieving real-time performance (117139 FPS) on consumer-grade hardware. This work aligns with **OA3** "Increase Responsiveness/Speed of Extreme Data Analysis Algorithms" by delivering a scalable, high-speed solution for privacy-aware visual data processing in intelligent transportation and surveillance applications. Figure 17 depicts an example of license plate detection and blurring using the AUTH method.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 17. License plate detection and blurring.

4.2.9. Skeleton-based action recognition

SOTA (incl. TEMA M1-M18)

Skeleton-based action recognition extends traditional person detection by not only identifying individuals but also analysing their movements and behaviours in disaster scenarios. This enables real-time detection of distress signals, identification of injured or trapped individuals, and improved situational awareness for first responders. Additionally, it helps assess crowd movement during evacuations, optimising resource allocation. Previous SOTA skeleton-based action recognition methods often rely on a single representation of the skeleton sequence, which limits their ability to capture complex action features effectively [156, 157]. Graph Convolutional Networks (GCNs) have been widely used for modelling skeleton-based actions, but they primarily focus on spatial dependencies between joints while overlooking joint motion dynamics, and suffer from over-smoothing as the network depth increases [158]. Attention-based models attempt to highlight important joints and frames, but they can be biased and difficult to control due to the diversity of actions [159, 160]. Additionally, many approaches lack invariance to viewpoint, motion speed, and sequence length, leading to performance degradation when these factors vary [161].

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Advances beyond SOTA

Motivated by the gaps of previous SOTA methods, AUTH introduces IMDAR. The details of the proposed method have been published in the "Information Sciences" journal [162]:

> Kamel, C. Zhang, and I. Pitas, "Spatio-temporal invariant descriptors for skeleton-based human action recognition," Information Sciences, vol. 700, p. 121832, 2025.

IMDAR is a framework that captures skeleton motion dynamics using multiple spatio-temporal invariant representations rather than relying on a single feature extraction approach. By structuring skeleton sequences into spatio-temporal matrices and transforming them into image descriptors, the method ensures robustness to viewpoint, motion velocity, and action complexity. A Voting and Priority Score Selection (VPSS) algorithm further enhances classification accuracy by selecting the best prediction from multiple descriptors. The developed approach demonstrates significant accuracy improvements over SOTA models across multiple benchmark datasets, confirming its effectiveness in addressing the limitations of existing methods. IMDAR has been evaluated on the NTU-RGB + D60 [10], which is a large-scale human action recognition dataset with 60 action classes, containing 56,880 skeleton sequences. Among others it contains classes related to medical conditions such as sneeze/cough, neck pain and headache.

Figure 18 demonstrates the detection of actions such as hand-wave, sneeze/cough, neck pain, and headache that are particularly relevant to NDM as they serve as key indicators of distress, injury, or health conditions in disaster scenarios. Hand-waving can signal for help in emergency situations, aiding in victim detection during search-and-rescue missions. Detecting sneeze/cough, neck pain and headache can assist in identifying individuals suffering from injury, fatigue, or exposure to hazardous conditions such as smoke inhalation or dehydration. By integrating skeleton-based action recognition into NDM, responders can gain valuable real-time insights into the well-being of affected individuals, leading to more efficient and targeted emergency assistance.

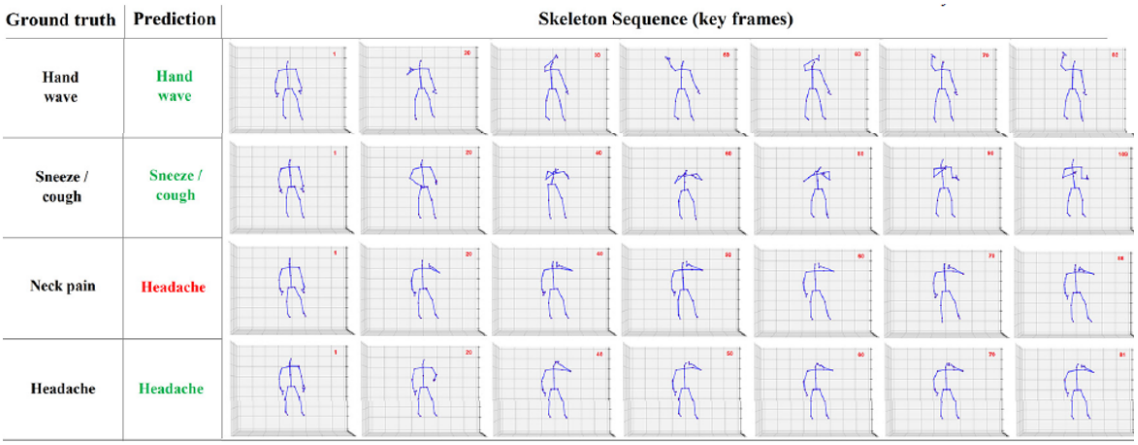


Figure 18. Predictions alongside the ground truth for samples from the testing set of the NTU-RGB + D60 dataset [10], where all test actions are captured with AUTH method (settings S=001, camera view C=001, performed by the performer P=003, and trial R=001). For each ground truth action, the key frames of the sequence are presented. Correctly classified actions are highlighted in green (despite high similarity with another action), while misclassified actions are highlighted in red (in cases of very high similarity with another action).

The comparison with the SOTA methods on NTU-RGB + D60 [10] for both Cross-Subject (C-Sub) and Cross-View (C-View) benchmarks [10] are shown in Table 10.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 10. Comparison of accuracy between IMDAR and the SOTA recognition methods on the NTU-RGB+D60 dataset [10] for Cross-Subject (C-Sub) and Cross-View (C-View) benchmarks (%).

Method	C-Sub	C-View
ST-GCN [163]	81.5	88.3
SR-TSL [164]	84.8	92.4
AS-GCN [165]	86.8	94.2
2s-AGCN [166]	88.5	95.1
AGC-LSTM [167]	89.2	95.0
RA-GCN [168]	87.3	93.6
4s-Shift-GCN [169]	90.7	96.5
Dynamic-GCN [170]	91.5	96.0
MST-GCN [156]	91.5	96.6
ST-TR [171]	90.3	96.3
Ta-CNN [172]	90.7	95.1
EfficientGCN [173]	92.1	96.1
ST-SLKA [174]	90.7	96.1
Action Capsules [159]	90.0	96.3
SHARL [175]	90.4	96.5
IMDAR	**92.8**	**96.8**

IMDAR demonstrates higher prediction accuracy on the C-Sub benchmark and performs even better than methods that combine both GCN and attention mechanisms. Specifically, it shows better performance than SHARL [175] by **2.4**% (C-Sub), which focuses on dynamic joint selection through reinforcement learning, while AUTH multiple invariant representations provide a comprehensive encoding of actions that can adapt to any dynamic change without the need for complex training frameworks.

4.2.10. Synthetic data generation

SOTA (incl. TEMA M1-M18)

The advances in generative AI over the past year have made it possible to address data scarcity in natural disaster management. Our previous approach leveraged the capabilities of Stable Diffusion XL (SDXL) [64], a state-of-the-art text-to-image diffusion model released in 2023, which produces high-resolution, photorealistic, and semantically rich images. SDXL represents a significant improvement in generating complex scenessuch as fires and floodsespecially when guided by descriptive prompts. In parallel, we employed zero-shot object detection models such as Grounding DINO [176], also introduced in 2023, which enable robust identification and segmentation of disaster-related visual elementssuch as smoke, flames, or waterwithout the need for fine-tuning on specific datasets. Together, these tools form the foundation for generating and annotating synthetic disaster imagery with a high degree of realism and contextual relevance.

Advances beyond SOTA

ATOS has also been advancing the generation of realistic visual images of natural disasters using SOTA diffusion models. These techniques enable the creation of synthetic images depicting

This project has received funding from the European Union – European Commission under grant agreement No 101093003

any geographic region under the impact of events such as floods or wildfires. In addition, we are developing methods to augment real-world aerial images captured by drones by overlaying synthetic elements like fire or water. This allows for the generation of realistic disaster scenarios enriched with geographical metadata, which can be seamlessly integrated into the TEMA platform for testing and validation, without the need for actual disaster occurrences. A comprehensive description of these methodologies and the outcomes produced will be provided in the document "D3.4 Report on AI Model Adaptability to Extreme Data".

4.2.11. Person re-identification

SOTA (incl. TEMA M1-M18)

Person Re-IDentification (ReID) has emerged as a critical component in natural disaster response systems, enabling the tracking and reunion of displaced individuals across disparate surveillance sources. The SOTA in 2024 reflected a growing shift toward lightweight, edge-compatible ReID models optimized for aerial imagery and constrained environments. Transformer-based methods and edge-enhanced networks [177] further contributed to improving performance under occlusion and lighting variability, common in disaster zones. These advancements highlighted the increasing importance of privacy-preserving, real-time ReID systems that operate efficiently on the edge, making them particularly relevant for humanitarian applications where data protection and operational agility are paramount. At the time the previous deliverable D3.1 was submitted, ATOS employed a ResNet50 model trained on the Market1501 dataset to extract unique embeddings for each person detection from the TEMA system.

Advances beyond SOTA

Recent efforts concentrated on optimising the system for high-resolution drone imagery. These aerial images offer a broader field of view and are especially valuable in disaster zones where ground-level visibility may be limited. However, their increased resolution presents significant computational challenges, particularly when aiming to achieve real-time performance. To address this limitation in D3.2, ATOS has focused on refining the solution to be deployed on an edge device achieving real time performance.

A key advantage of edge deployment in the TEMA project is the enhanced privacy it provides. By processing sensitive visual data such as identifiable images of individuals locally on the edge device inside a docker, the personal information is isolated from the system. This approach ensures that biometric data, like the feature embeddings used for ReID, remain within a secure and controlled environment. It aligns with data protection standards and ethical considerations, which are particularly important in humanitarian contexts where individuals rights and dignity must be preserved even during urgent operations.

4.3. Methods for satellite/SAR data analysis

4.3.1. Satellite-based flood detection and assessment

This component of TEMA (TFA-tech-08) is responsible for analysing Sentinel-1 GRD (radar) and Sentinel-2 MSI (multi-spectral) satellite images provided through the Copernicus Data Space Ecosystem (CDSE). It uses pre-trained Convolutional Neural Network (CNN) models such as U-Net, DeepLabv3+, and ResNet-based segmentation networks to delineate surface water bodies

This project has received funding from the European Union – European Commission under grant agreement No 101093003

in these images and distinguishes permanent from temporary flooded water. Based on the resulting georeferenced water maps, secondary information products like flood duration are derived. Beyond satellite-based flood detection that targets operational use and continuous monitoring, scientific studies and experiments were conducted regarding on-demand object detection in very high-resolution (<1m Ground Sample Distance (GSD)) orthoimages from drone surveys.

SOTA (incl. TEMA M1-M18)

Conventional rapid mapping methods for satellite-based flood mapping can be slow and labour-intensive [178]. Recent advances in deep learning and the availability of new large-scale remote sensing reference datasets have opened new possibilities for automated image analysis [179, 180, 181]. DLR deploys a modular processing chain for surface water monitoring that uses DNNs and enables automatic satellite data search, pre-processing, analysis, and dissemination [182]. Complementary to large-scale satellite image analysis, object detection in very high-resolution drone images can provide insights into where people are located and which assets are exposed during a disaster. Most common object detection methods include one stage detectors such as YOLO networks [183] and two stage detectors such as Region-based Convolutional Neural Networks (RCNN) [184]. Two stage detectors separate region proposal, classification and refinement of the location prediction, while one stage detectors conduct localisation and classification at once. Therefore, YOLO networks are reported to be faster compared to the potentially more accurate RCNNs.

Advances beyond SOTA

Satellite-based flood detection

Within TEMA, DLR further developed its flood monitoring methods, expanded existing reference datasets [185] and trained, validated and tested various DNN models for water segmentation (see D3.1). The best performing models for different satellite sensors have been implemented into DLRs fully automated processing chain [182]. Recent advances include, the deployment of the chain on a Kubernetes cluster within the DLR infrastructure for faster processing and the dissemination of results via Open Geospatial Consortium (OGC) Web Map Service (WMS) and SpatioTemporal Asset Catalog (STAC). The STAC Application Programming Interface (API) has been setup to enable queries by location, time and metadata properties, which allows the user to retrieve the relevant water masks for a given flood disaster. Public access to the API ensures seamless integration of products into the TEMA platform. A joint publication of DLR and PLUS/IT:U on "Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods" shows the added value of the TFA-tech-08 results to quickly identify disaster hotspots in data-scarce situations [186]:

> M. Wieland, S. Schmidt, B. Resch, A. Abecker, and S. Martinis, "Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods", Natural Hazards, 2025.

Speed of Sentinel-1 data acquisition through DLR receiving station using direct broadcast

DLR-DFD operates the Neustrelitz receiving station as part of the DLR Copernicus Collaborative Ground Segment. Strategically located in Europe and within the Sentinel-1 core ground stations reception area, Neustrelitz enables direct downlink of Sentinel-1 data. Downlink reception planning is fully automated based on user requests and European Space Agency (ESA)s Station

This project has received funding from the European Union – European Commission under grant agreement No 101093003

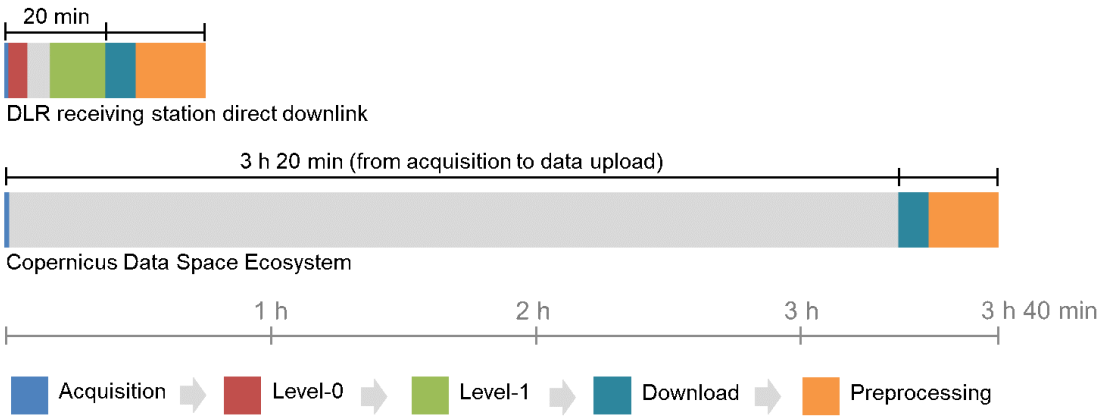


Figure 19. Comparison of the time delay from image acquisition by the Sentinel-1 satellite until availability for downstream analysis between DLR receiving station (via direct downlink) and Copernicus Data Space Ecosystem. Reported times are averaged across 30 Sentinel-1 images.

Downlink Plan. Sentinel-1 Level-1 product generation follows ESAs Payload Data Ground Segment (PDGS) Instrument Processing Facility (IPF) [187]. For satellite-based flood mapping with Sentinel-1, TFA-tech-08 uses the Level-1 Ground Range Detected product, which provides multi-looked, ground-projected data. To evaluate the speed improvements for data acquisition, we have set-up an experiment to measure the time delay from image acquisition by the Sentinel-1 satellite until availability for downstream analysis. Specifically, we compare times between data accessed through the DLR receiving station (via direct downlink) and via the publicly available Copernicus Data Space Ecosystem. A total of 30 Sentinel-1 images have been acquired over a pre-defined area of interest in Germany and times have been recorded for each of them. For a representative estimate, reported times are averaged across all 30 Sentinel-1 images for each data provider option.

The results in Figure 19 show that producing analysis-ready data via direct downlink at the DLR receiving station took approximately 40 minutes on average starting from the acquisition of the image by the satellite. Getting access to the analysis-ready image via the Copernicus Data Space Ecosystem took significantly longer with approximately 220 minutes on average. Therefore, our experiments indicate that the use of direct downlink could potentially reduce the time between sensing and availability of Sentinel-1 satellite images for downstream analysis by a factor of 5 compared to the publicly available Copernicus Data Space Ecosystem. **OC1 "Reduce latency in NDM"** has therefore been successfully addressed within the scope of TEMA.

Object detection in very high-resolution (<1m GSD) orthoimages

We compared the widely used FasterRCNN and YOLOv5 architectures to identify their potential for live-mapping in the context of rapid disaster response. We evaluated their performance on aerial and satellite images with respect to accuracy, processing speed and generalisation ability. For training, validation and test of the object detectors, we used two datasets: (i) the global xView benchmark dataset for object detection in satellite images [188] and (ii) a custom dataset that combines aerial images and bounding boxes of buildings in Germany. We reclassified the xView dataset and extracted only images that cover our classes of interest (buildings and vehicles). Images from both datasets were sliced into 512x512 pixel tiles for model training and validation.

YOLOv5 models are available at different complexity ranging from “nano” with 1.9M trainable

This project has received funding from the European Union – European Commission under grant agreement No 101093003

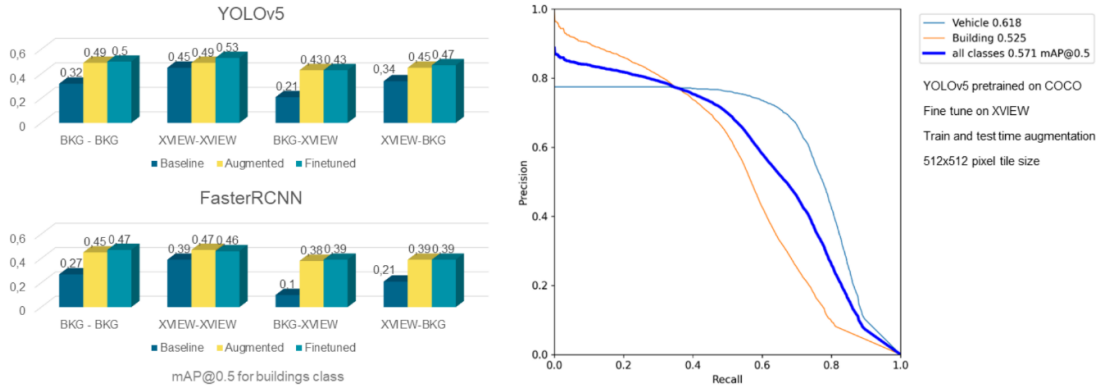


Figure 20. Accuracy assessment of trained object detectors. Left: Comparison of FasterRCNN models with YOLOv5l models for building class under varying data scenarios. Right: Precision-Recall-Curve of the final YOLOv5l model.

parameters to "extra-large" models with 140M parameters [189]. For our application YOLOv5l (with 46M parameters) offered a good trade-off between inference speed and accuracy as reported by the authors of the software. Similarly, we chose a Resnet50 backbone for the FasterRCNN model. We trained both YOLOv5l and FasterRCNN with a Stochastic Gradient Descent (SDG) optimizer, learning rate of 1e-3, momentum of 0.7 and weight decay of 2e-4, which have been obtained through initial experiments. We trained in batches of 8 for 100 epochs, while the best model state (determined by the lowest validation loss) was stored for evaluation. We tested the influence of data augmentation on the detection performance. Specifically, we applied with equal probability random brightness, contrast, scale and image flipping. We further evaluated different training strategies and explicitly compared how a model trained from scratch with randomly initialized weights (baseline) behaves compared to a model that is being fine-tuned on pre-trained weights on the MS COCO dataset.

Figure 20 shows the improved performance of YOLOv5l models compared to FasterRCNN models in all tested scenarios. Regardless of the model architecture used, we observed an added value of data augmentation with improvements in model accuracy of up to 0.28 mAP@0.5. Fine-tuning also had a consistently positive influence on model quality, even if the influence of this method was less than that of data augmentation. Models trained with augmented data also achieved higher accuracies when applied to data from other sensors than those on which they were trained. In our experiments, we emphasised that a trained model should be capable of generalising well across different imaging sensors and acquisition platforms. To this regard, we specifically tested the transfer between satellite imagery (XVIEW) and aerial imagery (BKG). The best model (YOLOv5l pretrained on COCO and fine-tuned on xView with train and test time augmentation) achieved a mAP@0.5 of 0.57 on the independent test-split of the reference dataset.

Figure 21 shows an example of model predictions on aerial images acquired by DLR's MACS camera system (Institute of Optical Sensor Systems) immediately after the Ahr valley flood on 2021-07-21. We can clearly observe car wrecks covered by mud either already aggregated on temporary parking areas or still scattered in muddy river parts (upper left). Emergency vehicles (upper right) and heavy machinery for debris removal (lower left) were successfully detected by our method. Clusters of vehicles parked along the borders of the most severely affected areas (lower right) could also be revealed.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Figure 21. Examples of detecting buildings (yellow) and vehicles (blue) in aerial images acquired by DLR's MACS camera system (Institute of Optical Sensor Systems) after the Ahr valley flood on 2021-07-21.

A good generalisation ability of the image analysis method is in any case essential to cope with highly varying data availability in disaster situations. Automated image processing routines together with pre-trained machine learning methods for image analysis can reduce the time between image acquisition and final product generation from several hours/days to just a few minutes. It therefore allows for a faster product delivery, for a higher analysis frequency and for a continuous monitoring of the situation.

Floods in Greece, Germany and Central Europe

Beyond the planned use-cases in TEMA, we tested TFA-tech-08 in three "live-scenarios" during the Thessaly floods in Greece September 2023 (see D3.1), the floods in Southern Germany June 2024 (Fig. 22) and the floods in Central Europe September 2024.

For the Central Europe floods, an area of 1,000,000+ km^2 across 8 countries was successfully monitored for 14 days. Real-time flood extent mapping produced 384 water masks from Sentinel-1 and Sentinel-2 satellite images. Moreover, permanent water bodies have been derived from timeseries analysis of 7,000+ Sentinel-2 water masks (2 years archive processing).

This project has received funding from the European Union – European Commission under grant agreement No 101093003

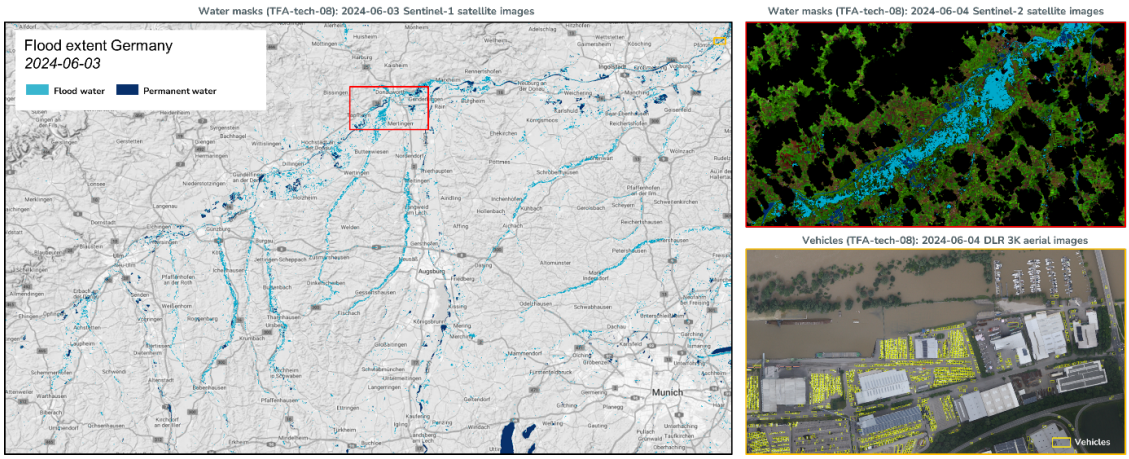


Figure 22. Flood situation over Southern Germany derived with TFA-tech-08 on 2024-06-03. Left: Flood extent from Sentinel-1 satellite images. Right-Top: Flood extent from Sentinel-2 satellite image. Right-Bottom: Object detection from DLR 3K aerial images (vehicles in yellow).

4.3.2. Satellite-based fire detection and assessment

The following component (TFA-tech-09) is responsible for the derivation of burnt area perimeters from mid- and high resolution optical imagery in Near Real-Time (NRT). The primarily used data sources are the Sentinel-3 OLCI and the Aqua/Terra MODIS sensors, alongside with Sentinel-2 MSI. An automated processing chain allows updates of the analysis results with every mid-resolution satellite overpass. This is implemented in a continental-scale monitoring system, featuring multiple updates of the wildfire situation throughout Europe per day. Besides the fire perimeters, the results include relevant information like the burn severity and the post-fire Normalized Difference Vegetation Index (NDVI).

SOTA (incl. TEMA M1-M18)

Despite the intensive development and advances considering the precision and applicability of global burnt area (BA) products [190], they still suffer from significant accuracy issues in many situations [191]. Boschetti et al. [192] validated the NASA MCD64A1 product with globally distributed Landsat image pairs, reporting a global Commission Error (CE) of 40.2% and a Omission Error (OE) and 72.6%, respectively. In addition, none of the global Burnt Area (BA) datasets available at this point can provide BA perimeters in NRT. For example, NASA's MCD64A1v061 algorithm maps BA using a time series-based approach retrospectively two months after the daily observations of MODIS [193].

Advances beyond SOTA

To address the prevailing accuracy issues of global BA products and the lack of available NRT information, we developed an NRT deep learning-based mapping procedure based on mid-resolution optical imagery and active fire data. It combines a novel superpixel-based segmentation technique and a GCN model to produce daily NRT BA perimeters.

To highlight the added value of the deep learning model within our mapping framework, we evaluated it against a preceding rule-based implementation of the DLR BA mapping algorithm

This project has received funding from the European Union – European Commission under grant agreement No 101093003

(DLRBAv1NTC) [28] and established machine learning models (Random Forest, LightGBM). We compared our Sentinel-3-based NRT (DLRBAv2NRT) and a monthly refined Non-Time-Critical (NTC) composite (DLRBAv2NTC) product against the established NTC MCD64A1v061 and the recently developed CGLBA31nrt product.

We conducted extensive evaluations for one of the most recent significant wildfire seasons in Greece in 2023. We also applied the DLRBAv2 algorithm on a variety of available mid- (Sentinel-3 OLCI, Terra MODIS, Suomi-NPP VIIRS) to high-resolution (Sentinel-2 MultiSpectral Instrument (MSI), Landsat 9 Operational Land Imager (OLI), Environmental Mapping and Analysis Program (EnMAP) HyperSpectral Imager (HSI)) optical imagery to produce NRT BA perimeters.

Figure 23 displays the derived BA perimeters of the four tested BA products for a fire in Rhodes, Greece, in July 2023.

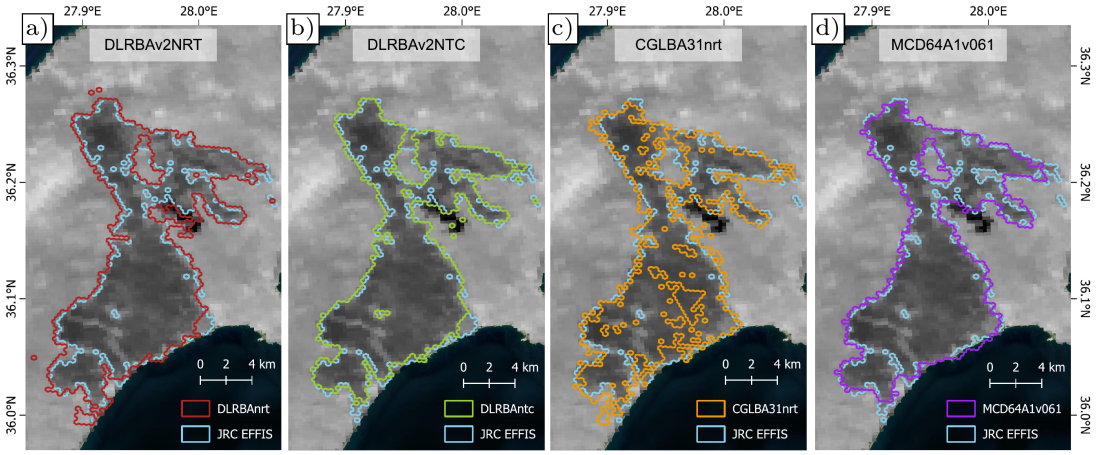


Figure 23. BA perimeter of a wildfire in Rhodes / Greece, July 2023, as mapped by the (a) DLRBAv2NRT, (b) DLRBAv2NTC, (c) CGLBA31nrt and (d) MCD64A1v061 product. Perimeters are displayed with the outlines of the burnt H3 cells. Background: Sentinel-3B band 17 (NIR) from 07/31/2023.

Product	Avg. IoU	Avg. F1-Score	Avg. Availability
CGLBA31nrt	0.62	0.76	+1 day
DLRBAv2NRT	0.69	0.81	< 1 h
DLRBAv2NTC	0.71	0.83	+1 month
MCD64A1v061	0.67	0.80	+2 months

Table 11. Average accuracy metrics and temporal availability of BA products over multiple study regions.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

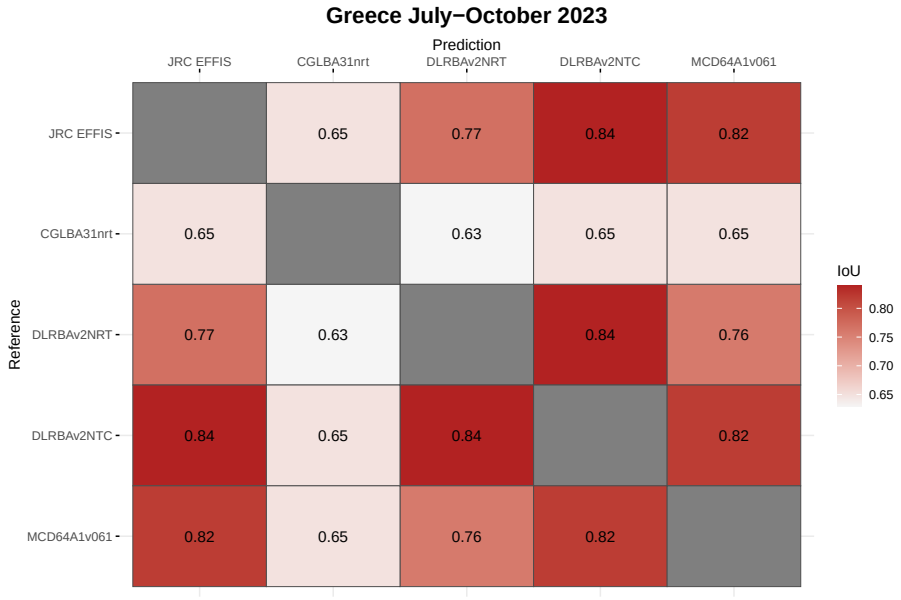


Figure 24. Intercomparison of BA products by IoU, regarding the Greece AOI.

Satellite-based burnt area derivation

The methodology described here is designed to address all the requirements of a large-scale BA monitoring system, with a focus on the timely availability of accurate BA information. It is a nested approach, embedding a deep learning step into a rule-based environment to derive BA based on thermal anomalies and optical satellite imagery.

Within TEMA, DLR further developed its burnt area monitoring methods and tested multiple rule-based, machine learning and deep learning approaches. The best performing approach for different satellite sensors has been implemented into DLRs fully automated processing chain. Recent advances include the deployment of the chain on a Kubernetes cluster within the DLR infrastructure for faster processing and the dissemination of results via OGC Web Feature Service (WFS). The Web Features API allows the users to query the available results by location, time and metadata properties. Public access to the API ensures seamless integration of products into the TEMA platform.

The proposed GCN model classified all BA in the 2023 Greece wildfire season with an IoU of 0.77 and F1-Score of 0.87, outperforming the Random Forest, LGBM and DLRBAv1NTC product (see Table 12 and Figure 25). The GCN model only produced a share of OE (6%) while achieving a satisfactory share of CE (19%). The results reported an improvement (+2% IoU and +1% F1 score) of the GCN model to the rule-based preceding DLRBAv1NTC product. As the DLRBAv1NTC product includes temporal refinement filters, Precision and Recall values were more balanced compared to the DLRBAv2NRT product, which tends to over-predict the BA more often but also has a significantly higher hit rate (Recall: 0.94). Only Random Forest and LGBM were able to detect BA with higher Recall (0.96). However, the GCN achieved a much higher Precision (0.81) in classifying BA compared to the Random Forest (0.71) and LGBM (0.72). Overall, with respect to KPIs, DLR achieved an improvement of 16% IoU for complete European coverage compared to NTC MCD64A1v06 which addresses successfully **OA2 "Increase accuracy of extreme data analysis algorithms"**. Through the use of a variety of

This project has received funding from the European Union – European Commission under grant agreement No 101093003

sensors the model implementation for TEMA reduces the latency between image acquisition and automatic 24/7 burnt area mapping by up to 2h compared to the manual resp. semi-automatic European Forest Fire Information System (EFFIS) approach (latencies of up to 6h on weekdays and >2 days on weekends). Accordingly, **OC1 "Reduce latency in NDM"** and **OA3 "Increase responsiveness/speed of extreme data analysis algorithms"** have been successfully addressed.

Model	IoU	F1-Score	Precision	Recall
DLRBAv2NRT GCN	0.77	0.87	0.81	0.94
DLRBAv2NRT Random Forest	0.69	0.82	0.71	0.96
DLRBAv2NRT LGBM	0.69	0.82	0.72	0.96
DLRBAv1NTC	0.75	0.86	0.87	0.85

Table 12. Accuracy metrics of tested classification models for Greece, 2023.

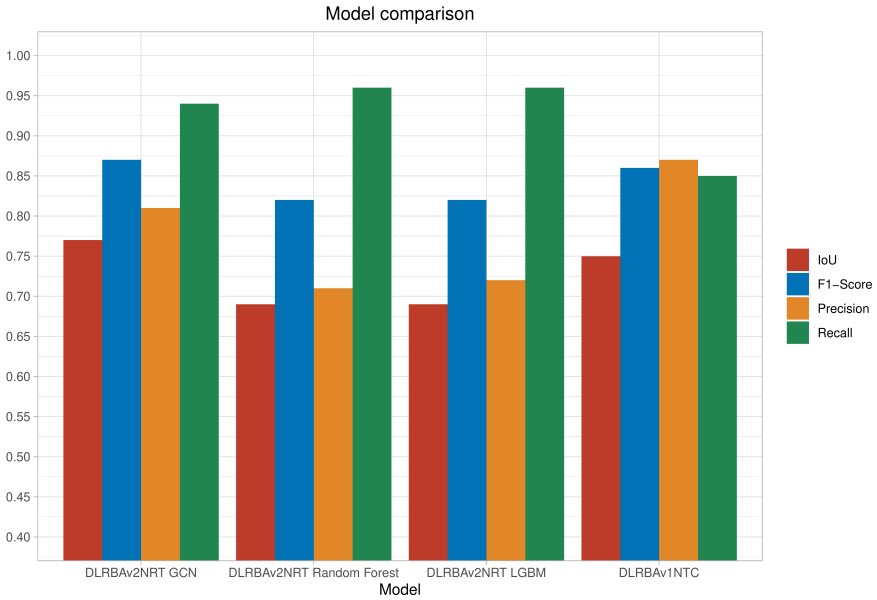


Figure 25. Accuracy metrics of tested models for Greece, 2023.

TEMA use cases

Figures 26 and 27 show BA mapping results for the TEMA use cases in Sardinia/Italy (Montiferru fire) and Kuhmo/Finland. Sentinel-2 post-event imagery was used for the BA derivation.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

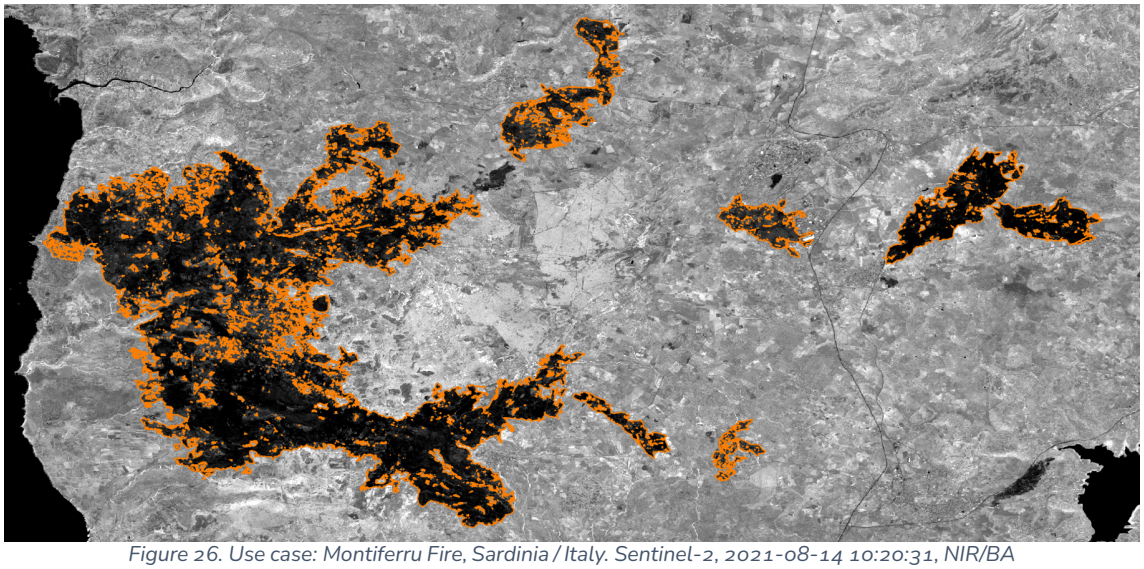


Figure 26. Use case: Montiferru Fire, Sardinia / Italy. Sentinel-2, 2021-08-14 10:20:31, NIR/BA

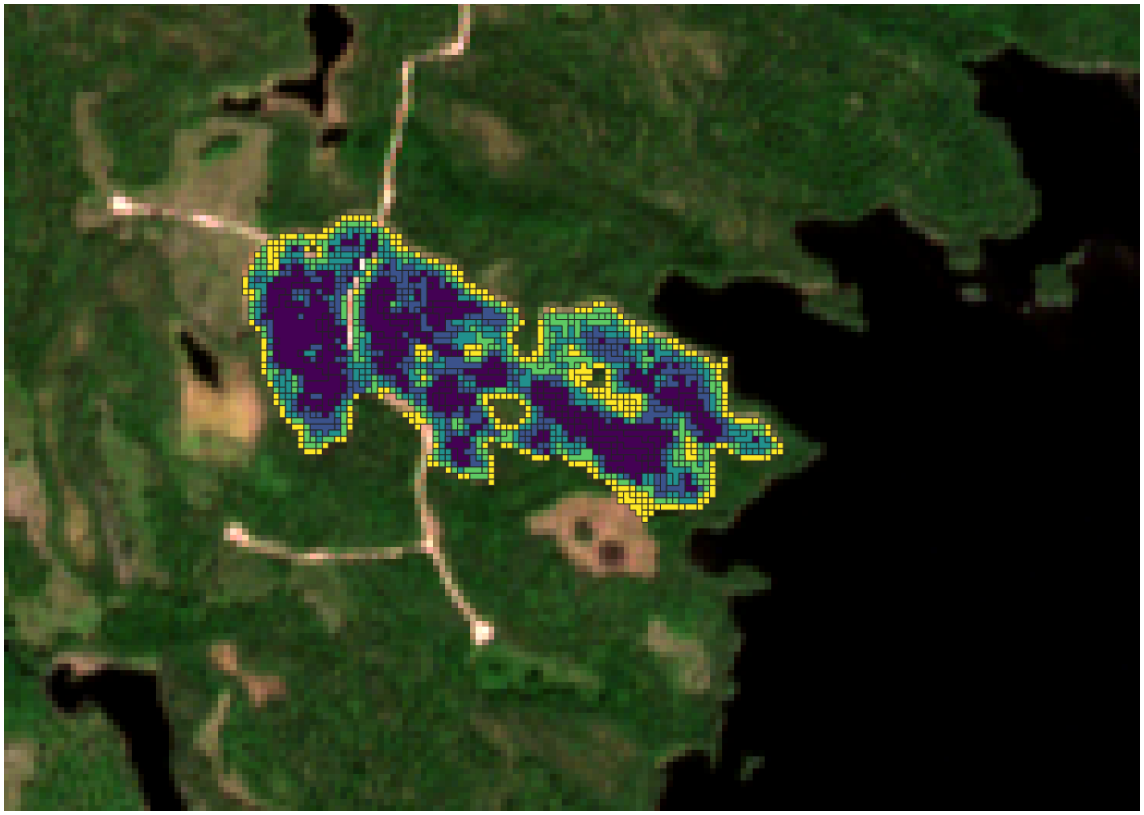


Figure 27. Use case: Fire near Kuhmo / Finland. Post-Fire NDVI (blue to yellow: lesser to higher vegetation fitness)

This project has received funding from the European Union – European Commission under grant agreement No 101093003

4.4. Analysis and Construction of 3D Smoke Concentration Maps

In D3.1, our 2D Finite Element Method (FEM) smoke model provided a foundational simulation of smoke dispersion dynamics under static conditions. In D3.2, this model has been extended and integrated into an operational context, enabling real-time smoke and wind mapping to support field-level decision-making. The main application of this module (PDM-tech-03) is for first responders and emergency response people to understand the dispersion of smoke from a forest fire. This is important since accurate information is crucial for evacuations and response planning; during forest fires, real-time wind and smoke maps can support fire-fighting activities on the ground [194].

SOTA (incl. TEMA M1-M18)

Existing smoke plume models are highly focused on model accuracy, in situations where many parameters, such as burn intensity and vegetation material, are known. We refer to [195] for an in-depth explanation of the state-of-the art algorithms for smoke plume dynamics. Their current drawbacks are in the norm model complexity and inadaptability to unknown inputs, both of which make them unsuitable for use in the TEMA platform, which requires real-time performance with some unknown inputs.

On the other hand, video games and computer graphics perform smoke model calculations in real time. These often suffer from many simplifying assumptions, since their priority is often artistic and a visually pleasing, "plausible" simulation is often preferred to a non-smoothly-rendered "realistic" simulation. For advances in computer graphics in smoke modelling, the review paper [196] provides more information. In a platform such as TEMA, where decisions will be made from the model's results, however, it is much preferred that the model reflects reality.

Both approaches to modelling share that, if they are physically accurate and not heuristically based or dependent on highly variable machine learning datasets, they solve some Partial Differential Equation (PDE) which describes the dynamics of the smoke. This is a computationally challenging task, particularly in three dimensions. Some simpler methods which work in two dimensions, such as finite differences, have lower numerical accuracy for the same number of computation points and run into issues when used in modelling real 3D data.

It is therefore a gold standard to have a model based on the use of finite elements, especially in cases where the domain of the problem is irregular, as is the case for the TEMA trials. One common method of doing so is using the FEniCS library [197]. DLR-KN in the past implemented some FEM-based code that modeled the spread of gas in an controlled environment for fire cases, which was mentioned in D3.1.

Advances beyond SOTA

In the last 12 months, the DLR-KN institute has focused on expanding the modelling capability of smoke and wind in the TEMA platform. We summarise this work in the following subsections.

3D Wind modelling

To integrate the topography into the scene, it is essential that the model uses terrain information in its calculations. DLR-KN has investigated two model types on the domains that take into

This project has received funding from the European Union – European Commission under grant agreement No 101093003

account real-time wind data and generate wind flow information. The first is a model which takes anemometer readings and extrapolates the results to the 3D domain (see Figure 28). The extrapolation takes into account the height over the surface at which every point is and uses the so-called log wind profile to obtain an estimate for the wind speed at this surface, assuming the same wind direction as measured on the ground. This model is expected to give accurate enough results for small domains but may pose some challenges when the wind sensors are far apart or the wind is uneven at higher altitudes.

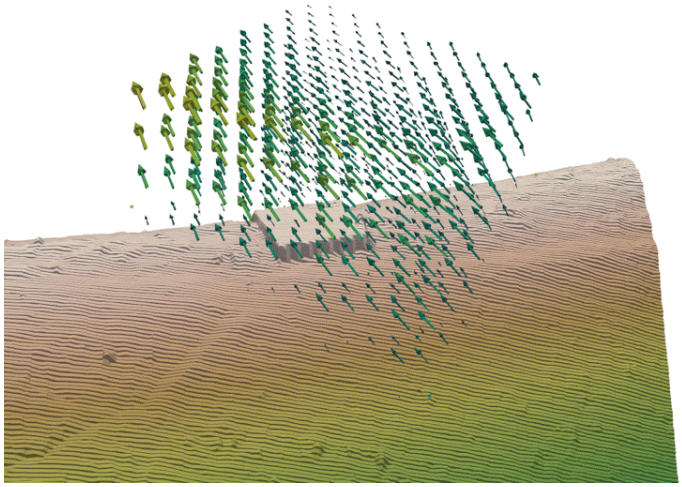


Figure 28. Interpolation-Based Wind Simulation in complex terrain. This wind field was generated from real data taken with wind sensors on the irregular topography of Vulcano, Italy in a measurement campaign.

The DLR also investigated another wind model, WindNinja [198], that uses more sophisticated CFD, to run simulations. This simulation generates 3D wind data which is informed by sensor readings on different elevation and models the interaction of this wind with uneven terrain. Figure 29 visualises this model's results.

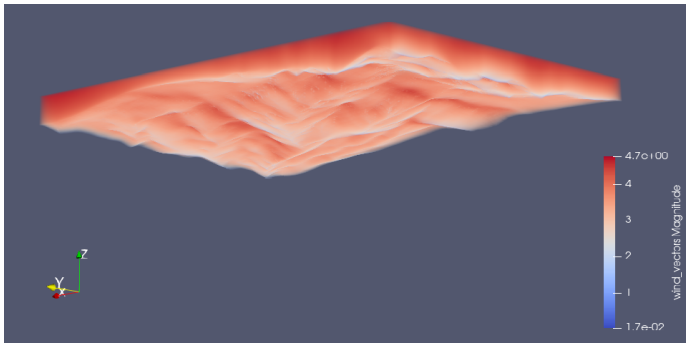


Figure 29. CFD-Informed Wind Simulation Based on WindNinja, visualized in Paraview. We see the wind speed visualized volumetrically over the Montiferru trial region. The mountainous terrain of the region is visible on the lower z bounds of the volume, and we can see that wind speed increases as expected the further one is from the wind surface.

Domain modelling

The FEM, requires a mesh of points over which the computations will be run, complete with a geometry of the whole domain. DLR-KN has made progress in defining a 3D mesh on which

This project has received funding from the European Union – European Commission under grant agreement No 101093003

wind and smoke models can run, as can be seen in Figure 30.

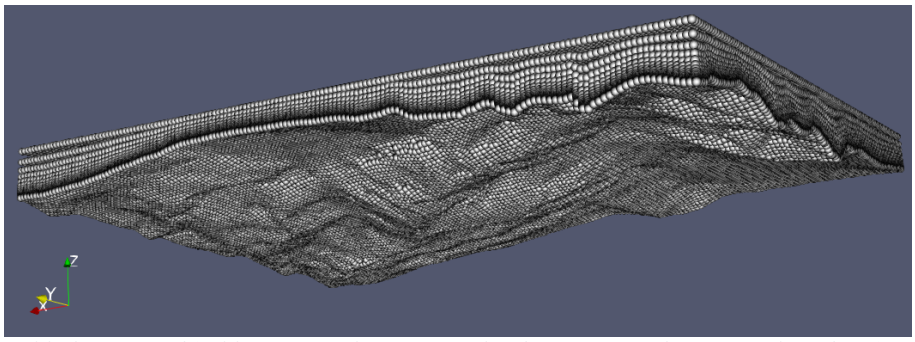


Figure 30. 3D Mesh generated for Montefiori trial region, visualized in Paraview. The region selected is of ca. 2km x 2km and has elevation difference of 500m within the selected points. Note the increased sampling frequency from bottom to the top in z. This is due to the calculation of wind.

The finite element models that DLR-KN uses to model smoke are informed from 3D wind information and currently run in two dimensions. They were implemented using the FEniCS library and are mentioned in the deliverables of WP4.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

5. Social media and text semantic analysis

5.1. Introduction

D3.2 builds upon D3.1 by introducing additional and improved models for semantic and emotion analysis of short texts and advancing multimodal analysis through the integration of spatiotemporal features and image-language models. This task involves IT:U, AUTH, DLR, ATOS and FHHI. IT:U leads the task, focusing on topic identification, multilingual handling, relevance classification and sentiment and emotion analysis. AUTH contributes by researching sentiment analysis for short texts, such as social media posts. ATOS investigates the use of Contrastive Language-Image Pre-training (CLIP) methodology to combine images with text snippets. These efforts are expected to enhance the capability to analyze geosocial media and news content accurately and efficiently, supporting the overall project goals.

Social media analysis has become an established technique for aiding efforts in disaster management, stretching over all four phases mitigation, preparedness, response, and recovery of the disaster management cycle [199, 200]. IT:U focused on topic modeling, multilingual handling, and post relevance classification, leading to the development of the JSTTS model and few-shot relevance classifiers. AUTH and IT:U introduced novel frameworks for short-text sentiment and emotion analysis, including the graph-based TMV approach, consensus-based labelling strategies, and Aspect-based Emotion Analysis (ABEA), which offers fine-grained emotional interpretation. ATOS evaluated contrastive image-language models for tasks such as image captioning and visual question answering, contributing to the multimodal understanding of disaster scenarios. Together, these developments directly support TEMA Objectives OA2 and OA3 by enhancing both the accuracy and responsiveness of extreme data analysis workflows.

5.2. Semantic topic modelling and analysis

Semantic analysis methods, including topic modelling and text classification, are essential for extracting meaningful information from large volumes of social media data. These techniques support situational awareness in disaster management by identifying relevant content, uncovering prevalent themes, and prioritizing actionable information.

5.2.1. Topic modelling

Topic modelling is a method for the unsupervised discovery of latent semantic topics within large collections of documents like textual social media posts. More generally, it can be described as a method of content analysis for text which involves assigning one or more "topic codes" to each document. Topic modelling can help increase situational awareness during disaster response and recovery by providing an overview of the topics present on social media [201]. It is also frequently used as a processing step for social media analysis in NDM, *e.g.*for the discovery of spatio-temporal topic clusters or hotspot analysis [202, 203]

This project has received funding from the European Union – European Commission under grant agreement No 101093003

SOTA (incl. TEMA M1-M18)

Traditional topic modelling has been dominated by probabilistic approaches such as Latent Dirichlet Allocation (LDA) [204], which infer semantic topics based on word co-occurrence under a bag-of-words assumption. While LDA has been extended to incorporate additional attributes such as sentiment, time, and user behaviour, it neglects word order, struggles with multilingual content, and performs poorly on noisy, short texts typical of social media data. More recent embedding-based models such as the Embedded Topic Model (ETM) [205], Top2Vec [206], and BERTopic [207] leverage pre-trained language models and clustering of semantic embeddings to improve topic coherence and interpretability. However, these approaches are limited to semantic information and fail to integrate other important dimensions like sentiment, spatial location, and time. Additionally, existing methods have not been systematically evaluated on disaster-related social media data, which often contains linguistic noise, emojis, slang, and multiple languages.

Within the first reporting period of TEMA, IT:U addressed these gaps by developing a family of multimodal topic modelling techniques. The Joint Topic-Sentiment (JTS) model extended clustering-based approaches by integrating sentiment as an additional modality, resulting in sentiment-associated topic clusters. JTS outperformed state-of-the-art methods like BERTopic, achieving a Topic Quality (TQ) score of 0.23 compared to BERTopics 0.12. Building on this, IT:U initiated the development of the Joint Spatio-Temporal Topic-Sentiment (JSTTS) model, incorporating spatial and temporal dimensions to enable geographically and temporally coherent topic-sentiment clustering. These advances represent a major step beyond existing methods, providing a comprehensive framework for analysing complex, multimodal disaster-related social media data. Results from JTS were published in [208], with further JSTTS developments contributing to T4.3 Information Fusion.

Advances beyond SOTA

The JSTTS model was the basis for a research paper presented at AGILE 2024 [209] and for a journal paper in the International Journal of Earth Observation and Geoinformation [210]. We specifically evaluated our integrated JSTTS approach against a traditional topic modelling workflow followed by sentiment classification and spatio-temporal analysis across four use case data sets concerned with (1) the 2021 Western Germany floods, (2) the 2017 Hurricane Harvey in Texas, (3) the 2023 Chile wildfires and (4) the 2014 Napa earthquake. Overall, our JSTTS approach achieved an average TQ score of 0.15 while a traditional topic modelling workflow followed by spatial clustering scored only 0.04. The reason for this is that traditional topic modelling techniques do not account for geographic space, resulting in redundant topics when applied in a pipeline spatial or temporal analyses.

Traditionally, semantic topics are represented by keywords which can limit interpretability. However, understandable outputs are crucial in a NDM context. Therefore IT:U implemented a topic summarisation approach using the generative AI model Llama-2 which is capable of extracting a short topic label. In addition, the model was successfully used to extract explicit information relevant to emergency responders from topic clusters.

With respect to KPIs, IT:U achieved an overall increase in Topic Coherence (TC) of 14%, enhanced Topic Diversity (TD) by 4% and 11% higher Topic Quality (TQ) compared to the SOTA. **OA2 "Increase accuracy of extreme data analysis algorithms"** has therefore been successfully addressed within the scope of TEMA. The algorithm implementation for TEMA additionally utilises highly efficient semantic embedding models like all-MiniLM-L6-v2 [211], allowing for the

This project has received funding from the European Union – European Commission under grant agreement No 101093003

rapid computation of embeddings and topic modelling in real time. **OA3 "Increase responsiveness/speed of extreme data analysis algorithms"** could therefore also successfully addressed.

5.2.2. Multilinguality handling

Naturally, social media posts come in several languages. Within TEMA, we therefore heavily focused on improving the performance of our methods across multiple languages.

SOTA (incl. TEMA M1-M18)

Natural Language Processing (NLP) methods have historically been dominated by language-independent approaches like bag-of-words or by neural network-based models, especially LLMs. However, most of these approaches were primarily designed for English. Consequently, they often struggle with multilingual content due to differences in vocabulary, morphology and grammar across languages. To address these challenges, multilingual models such as mBERT [212], XLM-RoBERTa [213], and GPT-4o [214] have emerged, supporting cross-lingual tasks by mapping documents from different languages into a shared embedding space. These multilingual models enable techniques like BERTopic [207] for multilingual topic modelling and classification. Nevertheless, their application to noisy, multilingual social media data especially in disaster contexts remains underexplored.

In the first reporting period of TEMA, IT:U addressed this gap by developing the JTS model [208], a multilingual topic modelling approach capable of analysing social media posts in up to 100 languages. By clustering semantic and sentiment embeddings, JTS generates sentiment-associated topics that significantly enhance interpretability and relevance. The model achieved a TQ score of 0.23 (compared to 0.12 for BERTopic), increased sentiment classification accuracy from 0.35 to 0.72, and improved sentiment uniformity within clusters to 0.9 (vs. 0.56 baseline). These advances represent a key step towards OA2 "Increase accuracy of extreme data analysis algorithms" by delivering robust multilingual performance, particularly valuable for cross-border disaster management scenarios.

Advances beyond SOTA

In the second reporting period of TEMA, IT:U further improved the JTS model, resulting in the JSTTS model already mentioned in Sec. 5.2.1. It allows for the extraction of geographically delineated topic-sentiment clusters of social media posts. The model supports multilingual analysis through multilingual embedding models like the multilingual Universal Sentence Encoder [215]. We evaluated it on geo-referenced Twitter data from four disaster events: the 2014 Napa Earthquake, Hurricane Harvey (2017), the 2021 Ahr Valley floods, and the 2023 Chile wildfires. Consequently, our experiments covered the languages English, German and Spanish. The spatio-temporal results of our method are covered in D4.2 due to their ties with information fusion. In this section, we focus on the language processing results. To assess the effectiveness of JSTTS, we conducted a comparative evaluation against a traditional sequential workflow. The baseline approach sequentially applies sentiment classification, topic modelling using BERTopic, and spatial clustering. Unlike JSTTSs integrated framework, this method processes each modality independently, potentially compromising coherence across semantic, sentiment, temporal and spatial dimensions.

The results in Table 13 demonstrate JSTTSs superior performance over the sequential workflow. Across all four disaster scenarios, JSTTS consistently outperformed the baseline in both Semantic

This project has received funding from the European Union – European Commission under grant agreement No 101093003

TQ and SU. TQ scores ranged from 0.081 (Napa Earthquake, English-only data) to 0.191 (Hurricane Harvey, predominantly English), compared to much lower values in the sequential approach (0.030 and 0.042, respectively). For the Ahr Valley floods (German) and Chile wildfires (Spanish), JSTTS achieved high TQ scores of 0.165 and 0.142, substantially exceeding the sequential methods 0.029 and 0.034. Similarly, SU was consistently higher with JSTTS, ranging from 0.837 (Chile wildfires) to 0.926 (Hurricane Harvey). This contrasted with the sequential workflow, where SU varied between 0.603 and 0.832. These results underline JSTTS's ability to generate semantically coherent and sentimentally consistent clusters, even in multilingual settings. By jointly modelling topics and sentiments while integrating spatial and temporal features, JSTTS outperforms traditional sequential methodologies that rely on independent processing steps.

Dataset	JSTTS TQ	Sequential TQ	JSTTS SU	Sequential SU
Ahr Valley floods	**0.165**	0.029	**0.891**	0.810
Hurricane Harvey	**0.191**	0.042	**0.926**	0.832
Chile wildfires	**0.142**	0.034	**0.837**	0.673
Napa earthquake	**0.081**	0.030	**0.894**	0.603

Table 13. Comparison of Topic Quality (TQ) and Sentiment Uniformity (SU) scores between JSTTS and a sequential workflow across four disaster-related use cases. Best scores are highlighted in bold.

The results further strengthen the achievement of the TEMA KPIs regaring **OA2 "Increase accuracy of extreme data analysis algorithms"**. The JSTTS consistently outperformed the sequential BERTopic-based workflow, achieving up to five times higher TQ scores across several languages for local and geographically delineated topic-sentiment clusters. All results regarding JSTTS model including the evaluation across multiple languages is published in [210].

To further validate multilingual topic modelling for disaster-related social media posts, we also established a multilingual topic modelling pipeline. It followed a language-agnostic text preprocessing workflow, including normalisation, tokenisation, and cleaning. Sentence-level embeddings were generated using transformer-based embedding models such multilingual-use. For topic modelling, BERTopic [207] was employed. The performance was evaluated using standard topic modelling quality metrics: TC, TD, Davies-Bouldin (DB), and Silhouette Score. TC measures the relative co-occurrence of top topic words based on pairwise Pointwise Mutual Information (PMI), TD assesses keyword uniqueness across topics, while DB and Silhouette Score quantify intra-cluster similarity and inter-cluster separation. These metrics align with the current SOTA [216].

Table 14 summarises the evaluation results across four disaster-specific datasets and a combined dataset. The Rhodes floods (gr) dataset achieved the highest TC (0.819), suggesting strong semantic consistency within topics. The combined dataset (it, gr, de, en) achieved the highest TD (0.810), indicating broader thematic coverage at the cost of weaker cluster separation, reflected by its higher DB score (2.940) and lower Silhouette Score (0.049). The Germany floods (de, en) and Montiferru wildfire (it) datasets showed moderate coherence and relatively high DB scores, pointing to less distinct cluster structures. These findings underline the general trade-off between topic coherence and diversity in multilingual topic modelling, confirming BERTopics effectiveness in challenging multilingual, real-world disaster data scenarios.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 14. Evaluation of topic modelling quality across multilingual disaster datasets, using Silhouette Score, DB Index, TC, and TD.

Dataset	Silhouette	DB Score	TC	TD
2021 Montiferru wildfires (Twitter)	0.102	2.383	0.517	0.650
2024 Rhodes floods (Twitter)	0.132	1.680	0.819	0.600
2021 Germany floods (Mastodon)	0.042	2.275	0.514	0.667
2021 Germany floods (Twitter)	0.079	2.053	0.577	0.603
Combined dataset	0.049	2.940	0.568	0.810

5.2.3. Post relevance classification/assessment

Differentiating between relevant and and irrelevant information on social media is crucial during natural disasters. IT:U therefore intensified their research on relevance classification within the second reporting period, focusing on improving situational awareness and low-resource settings.

SOTA (incl. TEMA M1-M18)

Relevance classification of geosocial media posts in disaster contexts remains a significant challenge due to the unstructured, noisy nature of the data and the inherent linguistic diversity across regions. Early methods primarily relied on keyword-based filtering, which proved inadequate for handling semantic ambiguity, multilingualism, and diverse linguistic structures. As the field advanced, machine learning approaches emerged, including CNNs [217], GNNs that integrate textual, visual, and temporal information [218], and transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) [219]. These models typically classified posts as informative or non-informative or assigned them to predefined actionability categories. However, most of these approaches were trained on English-language data, raising concerns about their generalisability to morphologically complex languages and multilingual settings. In parallel, the understanding of relevance has evolved from binary classification toward more fine-grained, multi-level relevance schemes that assess a posts contribution to situational awareness and crisis management efforts [220, 221]. Despite these developments, relevance classification methods tailored to the multilingual and multimodal nature of social media data in disaster scenarios remain underexplored.

Within the first reporting period of TEMA (M1-M18), IT:U addressed these gaps by developing and evaluating a text-based machine learning pipeline for relevance classification of disaster-related tweets, specifically in a flood context. This included a comparative analysis of multiple classifiers: a fine-tuned BERT model, naïve Bayes, random forest, SVM, and a custom CNN. IT:Us fine-tuned BERT model consistently outperformed all other methods, achieving improvements of 39 percentage points in the Gaussian score and 10 - 19 percentage points in the F1 score over traditional approaches. Additionally, analysis of misclassifications revealed that errors generally occurred between semantically similar categories, highlighting the robustness of the models classification logic. These results directly contribute to OA2 "Increase accuracy of extreme data analysis algorithms" in TEMA by improving relevance detection in disaster-related social media data. The findings were published in [221].

This project has received funding from the European Union – European Commission under grant agreement No 101093003

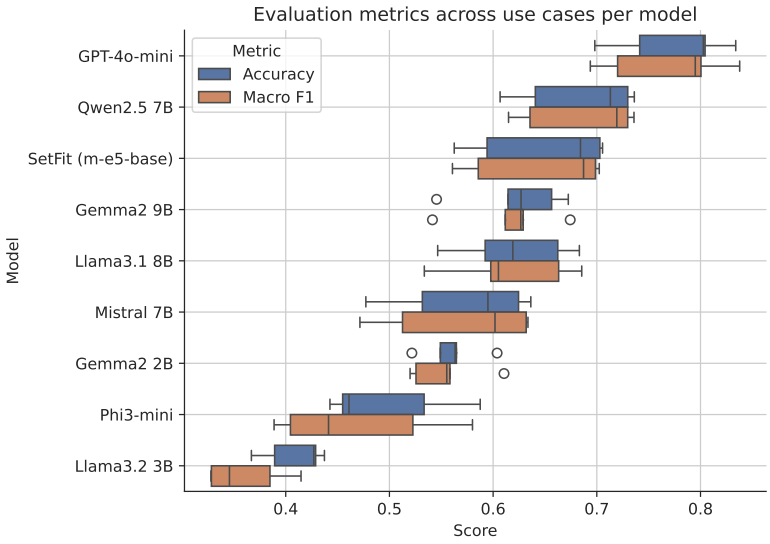


Figure 31. Distribution of accuracy and macro F1 scores across the five use case scenarios for each few-shot-learning model.

Advances beyond SOTA

Within TEMA, IT:U conducted an extensive evaluation of relevance classification strategies for geo-referenced social media posts in disaster response scenarios. The study focused on five major disaster events: the 2020 California wildfires, the 2021 Ahr Valley floods, the 2023 Chile wildfires, the 2023 Emilia-Romagna flood in Italy, and the 2023 Turkey/Syria earthquake. Naturally, our data spanned multiple languages, including English, German, Italian, Spanish, and Turkish. Given the urgent need for rapidly deployable, low-resource classification methods in time-critical disaster situations, we focused on Few-Shot Learning (FSL) approaches, which require minimal labelled data and no task-specific fine-tuning. We applied a three-class relevance scheme consisting of the classes *Related and relevant*, *Related but not relevant* and *Not related* and compiled an evaluation dataset of 4,746 posts using expert human annotation. For classification, we used few-shot prompting with eight Small Language Model (SLM), including GPT-4o-mini, Qwen2.5 7B, and Llama3.1 8B, and applied contrastive learning with SetFit fine-tuned on multilingual-e5-base. We classified posts in their original language to preserve semantic nuance and evaluated model performance using standard metrics such as accuracy, precision, recall, and macro F1-score.

The results are depicted in Fig. 31 and demonstrate the effectiveness of FSL approaches for rapid and accurate relevance classification in disaster contexts. GPT-4o-mini achieved the highest overall performance with an average macro F1-score of 0.77, followed by Qwen2.5 7B (0.69) and SetFit (0.65). Despite the diversity of disaster types and languages, GPT-4o-mini consistently delivered high accuracy, though classification performance was slightly lower for wildfire-related events, likely due to increased semantic ambiguity in post content. Notably, the multilingual-e5-base model, fine-tuned with SetFit, delivered competitive results while requiring significantly fewer computational resources. Our study focusing on few-shot learning for social media data in disaster management is the first of its kind and was accepted for the 7th International Conference on Advanced Research Methods and Analytics (CARMA) in July 2025.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

In an ongoing study, IT:U advances relevance classification of social media data by integrating spatial and temporal context alongside text-based features. The approach leverages a 13-dimensional feature vector comprising geographic and temporal proximity to disaster impact sites (derived from Earth Observation data), local densities and co-occurrences of disaster-related posts, as well as event-type and location encodings. Combined with transformer-based language models specialised for social media, such as TwHIN-BERT-base [222], the study systematically evaluates four multimodal integration strategies: feature concatenation, in-context learning, stacking, and partial stacking. Evaluations across 4,574 manually labelled posts from five global disaster scenarios show that while spatial and temporal features alone achieve strong performance (macro F1 = 0.713), and text-only classification with TwHIN-BERT-base reaches 0.779, the best results are achieved through partial stacking (macro F1 = 0.814). The proposed framework represents one of the first multilingual, context-aware approaches to relevance classification, contributing to the development of integrated GeoAI systems for disaster response

IT:U's research directly contributes to **OA2 "Increase accuracy of extreme data analysis algorithms"** within TEMA. With our enhanced meta learning approach, we were able to outperform our previous relevance classification model reported in [221] by 14% on the same use case, which were the 2021 Ahr Valley floods. Our Generative Pre-trained Transformer (GPT)-based few-shot learning method also outperformed our previous SOTA by 9% using only 5 samples per class. Consequently, our research also contributes to **OA3 "Increase responsiveness/speed of extreme data analysis algorithms"** by enabling rapid relevance classification in low-resource scenarios.

5.3. Sentiment analysis for short texts

Emotion analysis offers a more advanced and focused extension of sentiment analysis by providing deeper insights into human reactions. While traditional sentiment analysis typically classifies text into general categories like positive, negative, or neutral, emotion analysis particularly ABEA delves deeper by identifying specific emotional expressions tied to distinct aspects or topics. This approach significantly surpasses conventional sentiment analysis, which often overlooks nuanced emotional signals and tends to disregard aspect terms altogether.

5.3.1. Graph-based trustworthy majority voting

SOTA (incl. TEMA M1-M18)

The annotation of emotion in texts is a highly subjective task. Therefore, instead of relying on a single (and maybe biased) human annotator, it is commonly performed by multiple annotators. However, this leads to label inconsistency that needs to be addressed. To this end, emotion label aggregation methods have been proposed to account for annotator differences. Simple Majority Voting (SMV) is a widely-used baseline due to its simplicity, though it treats all annotators equally and thus does not consider annotator trustworthiness or biases [223]. Weighted Majority Voting (WMV) attempts to address this by adjusting annotator weights based on their general agreement with the majority, yet it neglects annotator pairwise agreement, limiting its effectiveness [224]. The Dawid-Skene model provides more accurate labels by probabilistically estimating annotator expertise but suffers from computational complexity [225]. Bayesian approaches extend Dawid-Skene by integrating prior knowledge, improving uncertainty management at the cost of computational overhead [226]. Multi-Annotator Competence Estimation (MACE) assesses annotator competence to filter out random labeling but struggles with subjective tasks due to difficulty defining annotator competence [227]. Deep Neural Networks with a

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Crowd Layer integrate annotator biases directly within a neural network architecture, requiring substantial data per annotator and potentially facing overfitting [228]. Lastly, Multi-Annotator Loss Modeling uses multi-task learning to separate noisy labels, enhancing robustness but at significant computational cost for large datasets [19]. Despite their advancements, these methods still face significant challenges in reliably aggregating subjective annotations.

Advances beyond SOTA

To address the label inconsistency problem, AUTH have developed a novel graph-based Trustworthy Majority Voting method, which is described in detail in a conference paper [229]:

> F. Augoustidis, P. Bassia and I. Pitas, "*Trustworthy Majority Voting for Labeling and Analyzing Multi-Annotator Text Sentiment Datasets*", European Signal Processing Conference (EUSIPCO), 2025.

The developed graph-based annotator ranking system consists of three steps: a) construction of an Annotator Agreement Graph (AAG), b) calculation of a Label Aggregation Score (LAS) for each annotator, c) the TMV scheme.

AAG $\mathcal{G} = \{V, E\}$ connects the annotators, which form the node set V, while edges E are formed between annotators if they have annotated at least a minimum number of T common data samples, which is a hyperparameter. The weight of each edge is set equal to the Cohen Kappa Score [230] between the two annotators, which quantifies the level of agreement between annotators on a set of jointly annotated text samples.

The LAS agreement score a_i for the $i-$th annotator is calculated as the average Cohen Kappa Score for each annotator and is normalized to [0,1], representing low or high annotator score agreement, respectively. Normalized LAS scores indicate annotator trustworthiness and are used to weight the annotations.

Finally, the developed TMV scheme operates as follows. For each data point and for each annotated label, a weighted aggregation score is calculated by summing all the weighted annotations. If a majority is formed, i.e., a unique (weighted) label score is higher than 50% of the sum of the LAS scores, the data is assigned that specific label. If no majority score is formed, the data point is simply discarded, as it is assumed that no reliable annotation exists for this data point.

To test the robustness of its method, AUTH introduced varying levels of label corruption to an existing trustworthy annotator, creating a mix of high and low-quality text sentiment annotations. The proposed method successfully decreased the trustworthiness of the corrupted annotator, thereby reducing its final contribution in the labeling aggregation process of the training text data. Furthermore, AUTH evaluated its approach under conditions of excessive label-level corruption, altering up to 50% of the total annotations. Models fine-tuned using the proposed AUTH aggregation technique demonstrate superior performance across all evaluation metrics.

Table 15 compares the balanced accuracy of the proposed TMV method against the previously SOTA Loss-Modeling method [19] across six Ekman sentiment classes [231] on the GoEmotions dataset [20].

It is clear that TMV consistently outperforms Loss-Modeling, demonstrating accuracy improvements ranging from **4.2**$\%$ (Fear) to as much as **13.4**$\%$ (Joy), surpassing the "**Sentiment analysis accuracy**" KPI of objective **OA2** "Increase accuracy of extreme data analysis algorithms". On average, TMV achieves an accuracy increase of approximately **7**$\%$, highlighting its effectiveness

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Table 15. Balanced accuracy comparison between TMV and the previous SOTA Loss-Modeling method [19] on GoEmotions dataset [20]. Higher values indicate better performance.

Sentiment Class	TMV	Loss-Modelling	% Increase
Anger	70%	67%	+4.4%
Disgust	68%	65%	+4.6%
Fear	73%	70%	+4.2%
Joy	76%	67%	+13.4%
Sadness	72%	68%	+5.8%
Surprise	76%	69%	+10.1%

in producing more reliable sentiment classifications by effectively accounting for annotator trustworthiness and mitigating noisy annotations.

5.3.2. Consensus-based labelling

SOTA (incl. TEMA M1-M18)

An important aspect of any type of model training is the quality of the available training data. This is particularly important in a rather subjective domain such as sentiment or even emotion analysis. However, traditionally such labelling is carried out by a single person or using crowd-sourcing platforms, which often leads to inconsistent or highly biased training datasets. So-called consensus-based labelling, which combines a multitude of annotators per data point, was therefore used to develop the models as part of TEMA. A data point is only included in the final training dataset based on a certain level of agreement between the assigned labels (e.g. 75%), a so-called inter-annotator agreement. To support this process, in-person labelling sessions were also conducted in which the annotators worked simultaneously but independently. After labelling, however, there was always the opportunity for exchange and discussion and thus potential adjustment of the label assigned.

In the first reporting period, consensus-based labelling was conducted for the training of an ABEA model, i.e. the identification of emotions related to specific words or phrases in an input text. A high-quality training dataset was created through a detailed labelling process, where human annotators identified emotions, marked relevant words, and classified them into *joy*, *anger*, *sadness*, or *fear*. A labelling guide ensured consistency, and consensus decisions refined the dataset, making it more specific than previous state-of-the-art approaches.

Advances beyond SOTA

Recent advances in LLMs have opened up new possibilities for the creation of training data. In particular, the speeding up and associated cost reduction of the traditionally slow process can be driven by this. As part of TEMA, state-of-the-art LLMs were tested for their application in domain-specific labelling. We found that this process already works well in principle, which is due to the LLMs' understanding of language. However, especially in less obvious cases (e.g. without clearly assignable keywords), deviating classifications from the human annotators were more frequent. As the data quality of training data is particularly important, no LLM-based training data was therefore used for TEMA.

The consensus-based labelling method was also used to generate a training dataset for

This project has received funding from the European Union – European Commission under grant agreement No 101093003

relevance classification. This domain is also very suitable, as a clear definition of relevance is very complex and labelling inaccuracies can therefore occur. Three experts from IT:U carried out the annotation simultaneously and discussed edge cases, which ensured the high quality of the created training data.

5.3.3. Aspect based emotion analysis

SOTA (incl. TEMA M1-M18)

Text emotion classification has evolved significantly, with lexicon-based, machine learning, and deep learning approaches offering different strengths. While lexicon-based methods provide simplicity [232], machine learning and deep learning models, including Naïve Bayes [233], SVMs [234], and LSTM models [235], have shown higher accuracy in capturing nuanced emotions. Additionally, transfer learning has become a valuable tool, especially for languages with limited labelled data [236]. Prior research has primarily focused on document- and sentence-level emotion analysis [237].

In the first reporting period, the Aspect-based Sentiment Analysis (ABSA) model GRadient hArmonized and CascadEd labeling (GRACE) by [238] was adapted for the task of ABEA. The aim of this model is to enhance social media analysis by accurately identifying and classifying emotional responses related to specific content in social media texts.

Advances beyond SOTA

Further hyperparameter optimisation using a manual grid-search approach was conducted to improve the relatively low model performance. With this, parameters such as batch size, and learning rate were optimised. Adjustments were also made to architectural elements such as shared layers between the Aspect Term Extraction (ATE) and Aspect Emotion Classification (AEC) branches. Fine-tuning also included dropout and weight decay adjustments, but some configurations led to overfitting or reduced performance, especially on external validation datasets. The model's performance was evaluated using F1-scores, and the highest performance was achieved with a configuration (50.8%). This final model utilized a batch size of 16, a learning rate of 2e-5, 6,000 epochs, 0.1 dropout rate, and 5 shared layers between the ATE and AEC branches, with fine-tuning applied to both branches in separate training steps. Cross-validation confirmed that the highest configuration outperformed the baseline by a 6-point improvement for joint ATE and AEC tasks. Since there is no comparable model for ABEA, this represents a large advancement of the SOTA. A pre-print describing the *EmoGRACE* model was published [239].

Additionally, we tested the usefulness of results from ABEA for disaster management in a case study on wildfires in California. The analysis found clear spatio-temporal patterns in emotional responses to wildfires. Areas near fire perimeters experienced significant increases in emotions like sadness, fear, and anger. Emotionally charged discussions peaked during key fire events, with anger and sadness dominating, while happiness emerged in relation to relief efforts which was presented as a Geospatial World Forum 2025 conference [240].

5.4. Spatial hot spot analysis

This project has received funding from the European Union – European Commission under grant agreement No 101093003

SOTA (incl. TEMA M1-M18)

Spatial hotspot detection identifies areas with a significantly higher-than-expected concentration of events based on statistical analysis. If the point density in a specific area deviates significantly from a random distribution, it is classified as a spatial hotspot [241]. Unlike heatmaps, which provide a continuous visual representation of density, hotspot detection methods assess statistical significance. Common techniques include spatial autocorrelation measures such as Morans I, Gearys C, and Gi* statistics [242]. Additionally, machine learning-based approaches have been explored for predicting hotspots [243].

The TEMA project applies spatial hotspot analysis to model the distribution of social media posts. To achieve this, various implementations were developed based on spatial aggregation of classified data, such as emotion-labelled tweets or disaster-related posts. Different aggregation methods, including hexagonal grids, were tested, with the Getis-Ord Gi* method identified as the most suitable. Additionally, the high temporal resolution of social media data enables spatiotemporal analysis, allowing for the detection of changes over time.

Advances beyond SOTA

In the TEMA project, hotspot analysis was employed to assess the potential for accelerating the satellite tasking process using social media data. Disaster-related Tweets were spatially aggregated and compared with a baseline from the previous year. A significant increase in the proportion of these posts was then categorised as a potential disaster event. The spatio-temporal distribution of these events was then compared to official alerts from Global Disaster Alert and Coordination System (GDACS) and burnt areas derived from Sentinel-3 imagery. The resulting study [244], which employed the workflow for two wildfire case studies, found that this method could be sensible in areas with higher population densities to speed up the acquisition of high-resolution remote sensing areas. This methodology, which is not part of standard remote sensing data retrieval and processing routines, thus advances the SOTA.

5.5. Contrastive image-language models

Vision Language Models (VLMs) represent an advanced evolution of AI, enabling the integration of visual and textual information for tasks such as image captioning and Visual Question Answering (VQA), in dynamic contexts like NDM.

SOTA (TEMA M1-M18)

Contrastive image-language models are relatively new, which is why there are not yet many studies using it. Image classification tasks, on the other hand, have received a lot of attention in the literature, also with regard to natural disasters. For the most part, CNN-based methodologies to identify images of floods posted on Twitter were developed [245, 219]. [246] propose a methodology based on a CNN and transfer learning to perform sentiment analysis on disaster-related imagery. [247] analyze which kind of image characteristics lead to higher user engagement on social media, using Google Cloud Vision to classify the image content. However, in the media-effective, on-going discourse about ChatGPT, BLIP-2 has also already been proposed as a tool for automatic question answering for visual content [248]. [249] compare the performance of BLIP-2 against other similar models such as OpenFlamingo, LLaVa or MiniGPT4. [250] find that BLIP-2 outperforms Flamingo and BLIP considerably for video

This project has received funding from the European Union – European Commission under grant agreement No 101093003

captioning. [251] propose a workflow to tag road segments from OpenStreetMap that employs BLIP-2, even though they still achieve a quite low accuracy.

At the project's inception, BLIP-2 [252] stood out as the only dependable image-to-text model, combining a CLIP image encoder with large language models such as Opt and Flan T5 to generate accurate captions and handle basic visual queries. Different variants, including blip2-opt-2.7b and blip2-flan-t5-xxl, showed comparable performance in producing short descriptions and answering general image-related questions. Shortly thereafter, OpenFlamingo [253] emerged as a more advanced alternative, integrating MosaicML Pretrained Transformer (MPT) models while still relying on a CLIP encoder. It demonstrated improved capabilities in addressing nuanced prompts and delivering richer visual outputs. LLaVA [254], built on Vicuna-7B and 13B, set a new benchmark for multimodal language models on 2023, significantly enhancing descriptive quality and offering advanced inference features, such as recognizing textual elements in images and managing follow-up interactions.

Advances beyond SOTA

ATOS has been actively studying and testing the latest developments in VLMs, which have significantly improved both the quality and depth of image captioning. These models are not only capable of generating accurate and contextually rich descriptions but also excel at answering complex questions based on visual inputs. This evolution is making VLMs increasingly relevant for real-world applications that require detailed scene understanding. Florence-2 [255], released in June 2024 by Microsoft, is a lightweight yet highly capable VLM. Despite its compact architecture offered in 0.2B and 0.7B parameter sizesit demonstrates strong zero-shot performance on a wide range of tasks, including image captioning, object detection, grounding, and segmentation. Its most notable strength lies in Visual Question Answering (VQA), achieving a competitive accuracy of 81.7% on the VQAv2 [256] benchmark. This balance of performance and efficiency makes Florence-2 one of the most scalable options in the VLM space. Phi-3.5-Vision [257], released in August 2024, is another significant contribution from Microsoft. It builds on the success of the Phi series by incorporating multimodal capabilities with a focus on high-level reasoning and visual understanding. With 4.2 billion parameters, it supports multi-frame image processing and is optimized for use in chat-based applications, making it ideal for dynamic, conversational AI systems. Phi-3.5-Vision is especially proficient in both VQA and image captioning, where it handles nuanced, multi-step reasoning questions with a degree of precision that places it close to larger models, despite its mid-size footprint. Molmo [258], developed by the Allen Institute for AI (Ai2) and released in September 2024, represents a powerful family of open-source multimodal models. The largest variant, with 72 billion parameters, rivals and even surpasses several proprietary models across various academic benchmarks. Molmo excels not only in image captioning and VQA but also in pointing and grounding tasks, which are critical for applications in robotics and virtual environments. It leads the performance chart with an impressive 86.5% accuracy on the VQAv2 benchmark, making it the strongest performer among the models evaluated. Qwen2.5-VL [259], released by Alibaba in February 2025, is a versatile model tailored for a broad spectrum of vision-language tasks. From image captioning and Optical Character Recognition (OCR) to interactive, dialogue-based VQA, Qwen2.5-VL demonstrates strong generalization and robustness. The 7B "Instruct" version achieves a 75.59% accuracy on VQAv2, showing that it balances capability and size effectively for more general-purpose multimodal applications. Compared to the previously studied models: BLIP2 (59.0%), OpenFlamingo (50.3%), and LLaVA-1.6 (81.8%), the latest VLMs especially Florence-2 and Molmo offer superior performance-to-size efficiency on the VQAv2 benchmark.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

This progress reflects not just scaling, but more effective training techniques and task alignment. After careful evaluation, ATOS selected Florence-2 as the most suitable VLM for integration into the TEMA platform. Its high accuracy in image captioning, combined with its compact size and efficient deployment footprint, makes it ideal for scalable, real-time applications.

The captions that can be generated from the image content sometimes also contain geographical information, such as place names. However, initial tests showed that these are often incorrect. A study [260] was therefore conducted to explicitly test the geocoding capabilities of vision language models for flood images. While they generally outperformed human annotators, the accuracy was still relatively low. In addition, a bias was found for Central Europe, where results were considerably better than for other world regions.

5.6. Explainability of language models

Within TEMA, IT:U developed and evaluated a multilingual disaster-related classification model of social media posts, leveraging Twitter-XLM-RoBERTa [261] fine-tuned on a curated, multilingual dataset comprising 179,391 disaster-related posts from CrisisLex [262], translated into five languages. Subsequently, the model was fine-tuned further for floods and wildfires using active learning. It demonstrated high classification performance, achieving an accuracy of 0.89 and a macro F1 score of 0.88 on a multilingual test set in Spanish, German and English. The model was published on HuggingFace alongside a conference paper presented at IntelliSys 2024 [263]. It is the **basis of several processing algorithms within TEMA** like hotspot detection and has been successfully used as a more accurate alternative to keyword filtering in recent studies [244]. The model further determines which posts are visualised within the TEMA system in T6.2. Given its key role within the project and its high performance, the model provides a robust foundation for applying XAI techniques. In a **collaboration of IT:U and FHHI**, we aimed to better understand its decision-making processes and to ensure trustworthiness in operational disaster response scenarios.

SOTA (incl. TEMA M1-M18)

Despite the growing adoption of large multilingual language models such as XLM-RoBERTa for disaster-related social media classification, their black-box nature remains a critical barrier to deployment in high-stakes environments like crisis response. While post-hoc explainability methods, particularly attention-based and gradient-based techniques like LRP [238, 200], they often fail to provide token-level insights that align with human-understandable disaster taxonomies. Moreover, few studies systematically explore how latent decision patterns evolve across transformer layers or how semantic coherence varies between activation and attribution-based representations. This lack of fine-grained interpretability limits trust in automated classifications, especially in multilingual and operational contexts.

Advances beyond SOTA

To achieve deeper explainability, we incorporate AttnLRP into the classification pipeline. It helps us track the effect of input tokens on the model's decision by propagating the class score in the backward direction through the model's layers. In practice AttnLRP is implemented using a modified gradient backpropagation pass, and token-level relevances are obtained using automatic differentiation. Subsequently, token-level relevance scores are normalized and displayed as heatmaps, explaining why a given decision was made in the classification. In parallel, to analyze

This project has received funding from the European Union – European Commission under grant agreement No 101093003

the model's latent structure, we extracted relevance vectors (via AttnLRP) and activation vectors (layer output) from three representative layers: one intermediate layer, one output layer, and the final classification layer. These high-dimensional vectors were reduced using 2D Principal Component Analysis (PCA) projection and grouped into clusters according to the disaster dataset from which those the tweets were extracted. This enabled the discovery of latent semantic groupings in disaster-related posts ($LABEL_1$) and non-disaster posts ($LABEL_0$). The relevance and activation-based clustering were examined individually and conditionally based on the predicted label. This exposed the development of meaningful clusters along semantics, such as fire, water, or emergency content. Discovering these meaningful clusters provides valuable insights into the model's decision-making process. The relevances of the tokens was calculated to rank the top tokens per cluster using the attribution score to bring up representative words corresponding to the taxonomies of human disasters. The pipeline was applied to single datasets (e.g., Chile wildfires, Germany floods) and multilingual aggregate corpora, illustrating the framework's universality across languages and disaster types. We also constructed relevance flow plots, further increasing traceability, to illustrate the progression of attribution values through the layers of the models. These plots exclude the last layer of classification to bring to the fore the interpretability of the decision steps in the middle of the transformer architecture.

Integrating AttnLRP with the multilingual XLM-RoBERTa classifier produced transparent visualizations that explain how the model distinguishes between disaster-related ($LABEL_1$) and non-disaster-related ($LABEL_0$) tweets. Figures 32 and 33 revealed that relevance-based embeddings form more coherent and semantically meaningful clusters than activations, especially for $LABEL_1$. This suggests that relevance vectors better capture the key factors driving the models decisions.

In the transformer layers, most prominently the final classifier, dense layer $LABEL_1$ tweets clustered closer to each other. These clusters consistently included words such as "flood," "terremoto," "evacuate," and "emergency," indicating that the model's internal reasoning aligns with human reasoning. While earlier layers showed greater label overlap, the relevance traces across layers highlighted a clear trend: deeper layers increasingly disentangle semantic patterns related to disaster classification.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

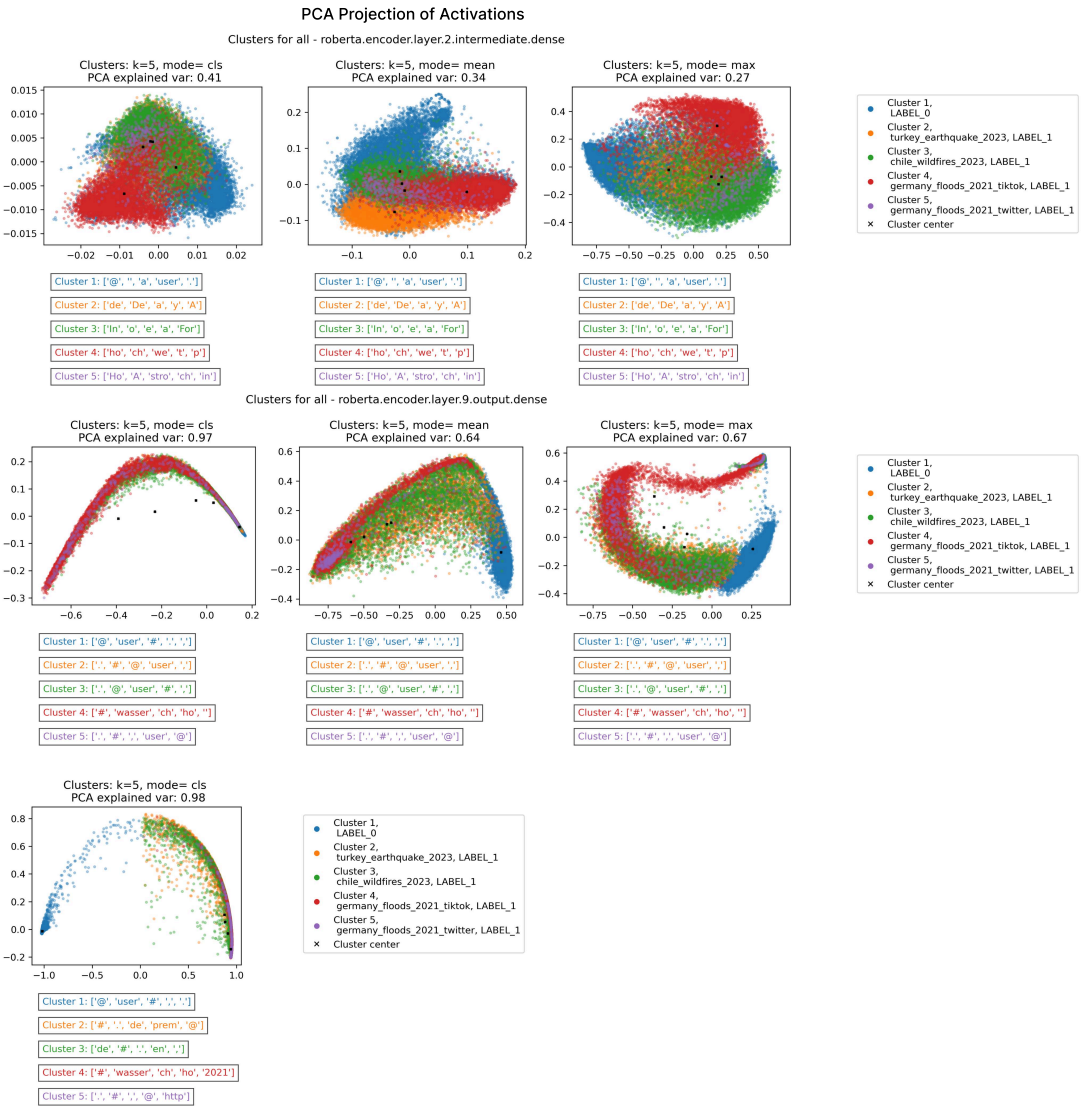


Figure 32. PCA Projection of Activations

This project has received funding from the European Union – European Commission under grant agreement No 101093003

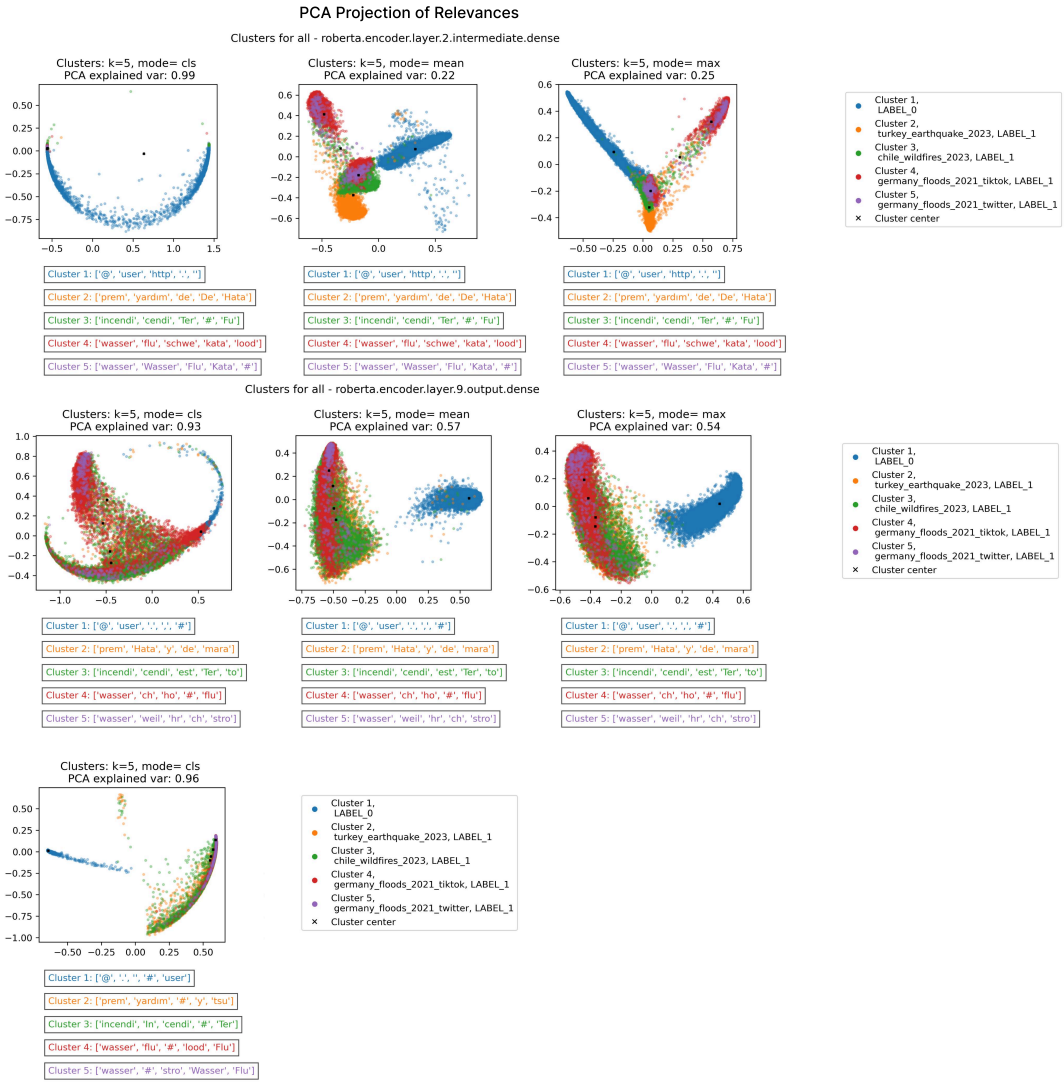


Figure 33. PCA Projection of Relevances

This project has received funding from the European Union – European Commission under grant agreement No 101093003

6. Conclusion

Deliverable D3.2 "Final report on algorithms for extreme data analytics" is the second deliverable of WP3 of the TEMA project. This document reports the final research results of Tasks T3.1 "Explainable and robust analytics", T3.2 "Real-time semantic visual analysis and remote sensing", T3.3 "Social media and text semantic analysis" between M19-M30. The main outputs of the research carried out, were **32 peer-reviewed publications**, that were accepted for academic journals and conferences and **8 technical reports** published as preprints. The developments are interlinked with WP4, WP5 and WP6. This document serves as a summary of the main research outputs and serves as a reference point for researchers summarizing the technical challenges of designing novel algorithms for extreme data analytics. As a public deliverable, it also supports the dissemination of results to the broader scientific community. Importantly, **all KPIs, objectives, and target values defined under WP3**, including improvements in algorithm accuracy, trustworthiness, responsiveness and speed, have been successfully addressed. These include significant advancements in explainability, semantic segmentation, object detection, semantic topic and sentiment analysis, relevance classification and emotion analysis of short texts. Performance improvements consistently **exceeded state-of-the-art baselines**, validating the technical effectiveness and scientific relevance of the approaches developed within TEMA.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

References

[1] X. Chen, B. Hopkins, H. Wang, L. ONeill, F. Afghah, A. Razi, P. Fulé, J. Coen, E. Rowell, and A. Watts, "Wildland fire detection and monitoring using a drone-collected rgb/ir image dataset," *IEEE Access*, vol. 10, pp. 121301–121317, 2022.

[2] Y. Sun, W. Zuo, and M. Liu, "Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2576–2583, 2019.

[3] S. Dong, W. Zhou, C. Xu, and W. Yan, "Egfnet: Edge-aware guidance fusion network for rgb–thermal urban scene parsing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 657–669, 2023.

[4] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku, and T. Harada, "Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5108–5115, IEEE, 2017.

[5] Z. Wan, Y. Wang, S. Yong, P. Zhang, S. Stepputtis, K. P. Sycara, and Y. Xie, "Sigma: Siamese mamba network for multi-modal semantic segmentation," *CoRR*, 2024.

[6] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19529–19539, 2023.

[7] C. Liu, Y. Zhang, Q. Chen, I. Pitas, and R. Fan, "These maps are made by propagation: Adapting deep stereo networks to road scenarios with decisive disparity diffusion," *IEEE Transactions on Image Processing*, vol. 34, pp. 1516–1528, 2025.

[8] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2024.

[9] Y. Xu, W. Yan, G. Yang, J. Luo, T. Li, and J. He, "Centerface: joint face detection and alignment using face as point," *Scientific Programming*, vol. 2020, no. 1, p. 7845384, 2020.

[10] A. Shahroudy, J. Liu, T.-T. Ng, and G. Wang, "Ntu rgb+d: A large scale dataset for 3d human activity analysis," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1010–1019, 2016.

[11] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.

[12] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *International Conference on Artificial Intelligence and Statistics*, 2011.

[13] S. Michalis and F. Kitsos, "Blaze fire classification segmentation dataset," June 2024.

[14] T. Toulouse, L. Rossi, A. Campana, T. Celik, and M. A. Akhloufi, "Computer vision for wildfire research: An evolving image dataset for processing and analysis," *Fire Safety Journal*, vol. 92, pp. 188–194, 2017.

[15] M. Gygli, H. Grabner, H. Riemenschneider, and L. V. Gool, "Creating summaries from user videos," in *Computer Vision – ECCV 2014*, pp. 505–520, Springer International Publishing, 2014.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[16] Y. Song, J. Vallmitjana, A. Stent, and A. Jaimes, "Tvsum: Summarizing web videos using titles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5179–5187, IEEE, 2015.

[17] S. D. Vito, E. Massera, M. Piga, L. Martinotto, and G. D. Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," *Sensors and Actuators B-chemical*, vol. 129, pp. 750–757, 2008.

[18] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7380–7399, 2021.

[19] U. Jinadu, J. Annan, S. Wen, and Y. Ding, "Loss modeling for multi-annotator datasets," *ArXiv*, vol. abs/2311.00619, 2023.

[20] D. Demszky, D. Movshovitz-Attias, J. Ko, A. S. Cowen, G. Nemade, and S. Ravi, "Goemotions: A dataset of fine-grained emotions," in *Annual Meeting of the Association for Computational Linguistics*, 2020.

[21] R. Tang, L. Liu, A. Pandey, Z. Jiang, G. Yang, K. Kumar, P. Stenetorp, J. Lin, and F. Ture, "What the DAAM: Interpreting stable diffusion using cross attention," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023.

[22] D. Papaioannou, V. Mygdalis, and I. Pitas, "Forest fire image classification through decentralized dnn inference," in *2024 IEEE International Conference on Image Processing Challenges and Workshops (ICIPCW)*, pp. 4134–4140, IEEE, 2024.

[23] D. Papaioannou, V. Mygdalis, and I. Pitas, "A decentralized sharding bft consensus approach, for efficient decentralized dnn inference classification," Apr. 2025.

[24] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

[25] A. Gerontopoulos, D. Papaioannou, C. Papaioannidis, and I. Pitas, "Real-time flood water segmentation with deep neural networks," Apr. 2025.

[26] P. Mentesidis, V. Mygdalis, and I. Pitas, "Improve real-time flood segmentation by encoding and distilling foreground information," May 2025.

[27] M. D. Tzimas, V. Mygdalis, and I. Pitas, "A weighting loss approach for transformer-based object detection," May 2025.

[28] M. Nolde, S. Plank, and T. Riedlinger, "An Adaptive and Extensible System for Satellite-Based, Large Scale Burnt Area Monitoring in Near-Real Time," *Remote Sensing*, vol. 12, no. 13, p. 2162, 2020.

[29] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," *PLOS ONE*, vol. 10, pp. 1–46, 07 2015.

[30] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 818–833, 2014.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[31] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," in *International Conference on Learning Representations (ICLR)*, 2014.

[32] A. Hedström, P. Bommer, K. K. Wickstrøm, W. Samek, S. Lapuschkin, and M. M. C. Höhne, "The Meta-Evaluation Problem in Explainable AI: Identifying Reliable Estimators with MetaQuantus," *Transactions on Machine Learning Research*, 2023.

[33] A. Hedström, L. Weber, S. Lapuschkin, and M. Höhne, "A Fresh Look at Sanity Checks for Saliency Maps," in *World Conference on Explainable Artificial Intelligence*, 2024.

[34] K. Dawoud, W. Samek, P. Eisert, S. Lapuschkin, and S. Bosse, "Human-Centered Evaluation of XAI Methods ," in *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 912–921, IEEE Computer Society, 2023.

[35] B. Kim, M. Wattenberg, J. Gilmer, C. J. Cai, J. Wexler, F. B. Viégas, and R. Sayres, "Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV).," in *ICML*, vol. 80, pp. 2673–2682, 2018.

[36] D. Bareeva, M. Dreyer, F. Pahde, W. Samek, and S. Lapuschkin, "Reactive Model Correction: Mitigating Harm to Task-Relevant Features via Conditional Bias Suppression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.

[37] M. Dreyer, F. Pahde, C. J. Anders, W. Samek, and S. Lapuschkin, "From hope to safety: Unlearning biases of deep models via gradient penalization in latent space," *Proceedings of the AAAI Conference on Artificial Intelligence*, no. 19, 2024.

[38] M. Dreyer, R. Achtibat, W. Samek, and S. Lapuschkin, "Understanding the (Extra-)Ordinary: Validating Deep Model Decisions with Prototypical Concept-based Explanations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.

[39] A. Gamaleldin, A. Atef, H. Saker, and A. Shaheen, "FIRE Dataset - Created by the NASA Space Apps Challenge 2018 and released via Kaggle," 2018.

[40] M. Dreyer, R. Achtibat, T. Wiegand, W. Samek, and S. Lapuschkin, "Revealing hidden context bias in segmentation and object detection through concept-specific explanations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3828–3838, 2023.

[41] M. Dreyer, E. Purelku, J. Vielhaben, W. Samek, and S. Lapuschkin, "PURE: Turning Polysemantic Neurons Into Pure Features by Identifying Relevant Circuits," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.

[42] F. Bley, S. Lapuschkin, W. Samek, and G. Montavon, "Explaining predictive uncertainty by exposing second-order effects," *Pattern Recognition*, vol. 160, p. 111171, Apr. 2025.

[43] J. Vielhaben, S. Lapuschkin, G. Montavon, and W. Samek, "Explainable AI for Time Series via Virtual Inspection Layers," *Pattern Recognition*, vol. 150, p. 110309, 2024.

[44] S. Becker, J. Vielhaben, M. Ackermann, K.-R. Müller, S. Lapuschkin, and W. Samek, "AudioMNIST: Exploring Explainable Artificial Intelligence for audio analysis on a simple benchmark," *Journal of the Franklin Institute*, vol. 361, no. 1, pp. 418–428, 2024.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[45] A. Frommholz, F. Seipel, S. Lapuschkin, W. Samek, and J. Vielhaben, "XAI-based Comparison of Audio Event Classifiers with different Input Representations," in *Proceedings of the 20th International Conference on Content-Based Multimedia Indexing*, p. 126132, 2023.

[46] C. Tinauer, A. Damulina, M. Sackl, M. Soellradl, R. Achtibat, M. Dreyer, F. Pahde, S. Lapuschkin, R. Schmidt, S. Ropele, W. Samek, and C. Langkammer, "Explainable Concept Mappings of MRI: Revealing the Mechanisms Underlying Deep Learning-Based Brain Disease Classification," in *Proceedings of the 2nd World Conference on Explainable Artificial Intelligence (XAI)*, 2024.

[47] G. Ümit Yolcu, T. Wiegand, W. Samek, and S. Lapuschkin, "DualView: Data Attribution from the Dual Perspective," *arXiv preprint arXiv:2402.12118*, 2024.

[48] F. Pahde, M. Dreyer, M. Weckbecker, L. Weber, C. J. Anders, T. Wiegand, W. Samek, and S. Lapuschkin, "Navigating neural space: Revisiting concept activation vectors to overcome directional divergence," in *The Thirteenth International Conference on Learning Representations ICLR*, 2025.

[49] E. Schnoor, M. Tiomoko, J. Said, S. Lapuschkin, and W. Samek, "Concept Activation Vectors from a Statistical Learning Perspective," *Poster presentation at the 7th Joint Statistical Meeting of the Deutsche Arbeitsgemeinschaft Statistik (DAGStat 2025)*, 2025.

[50] F. Pahde, T. Wiegand, S. Lapuschkin, and W. Samek, "Ensuring Medical AI Safety: Explainable AI-Driven Detection and Mitigation of Spurious Model Behaviorand Associated Data," *arXiv preprint arXiv:2501.13818*, 2025.

[51] J. Vielhaben, D. Bareeva, J. Berend, W. Samek, and N. Strodthoff, "Beyond Scalars: Concept-Based Alignment Analysis in Vision Transformers," in *The Thirteenth International Conference on Learning Representations ICLR Workshop - Workshop on Representational Alignment (Re-Align)*, 2025.

[52] E. Erogullari, S. Lapuschkin, W. Samek, and F. Pahde, "Post-Hoc Concept Disentanglement: From Correlated to Isolated Concept Representations," in *Explainable Artificial Intelligence, Third World Conference, xAI 2025*, 2025.

[53] L. Kopf, P. L. Bommer, A. Hedström, S. Lapuschkin, M. M.-C. Höhne, and K. Bykov, "CoSy: Evaluating Textual Explanations of Neurons," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 34656–34685, 2024.

[54] B. Puri, A. Jain, E. Golimblevskaia, P. Kahardipraja, T. Wiegand, W. Samek, and S. Lapuschkin, "FADE: Why Bad Descriptions Happen to Good Features," *arXiv preprint arXiv:2502.16994*, 2025.

[55] D. Bareeva, G. Ümit Yolcu, A. Hedström, N. Schmolenski, T. Wiegand, W. Samek, and S. Lapuschkin, "Quanda: An Interpretability Toolkit for Training Data Attribution Evaluation and Beyond," in *NeurIPS'24 Workshop on Attributing Model Behavior at Scale (ATTRIB)*, 2024.

[56] L. Arras, B. Puri, P. Kahardipraja, S. Lapuschkin, and W. Samek, "A Close Look at Decomposition-based XAI-Methods for Transformer Language Models," *arXiv preprint arXiv:2502.15886*, 2025.

[57] R. Achtibat, S. M. V. Hatefi, M. Dreyer, A. Jain, T. Wiegand, S. Lapuschkin, and W. Samek, "AttnLRP: Attention-Aware Layer-wise Relevance Propagation for Transformers," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2024.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[58] S. M. V. Hatefi, M. Dreyer, R. Achtibat, T. Wiegand, W. Samek, and S. Lapuschkin, "Pruning By Explaining Revisited: Optimizing Attribution Methods to Prune CNNs and Transformers," in *ECCV'24 Workshop on Explainable Computer Vision (eXCV)*, 2024.

[59] L. Weber, J. Berend, M. Weckbecker, A. Binder, T. Wiegand, W. Samek, and S. Lapuschkin, "Efficient and Flexible Neural Network Training through Layer-wise Feedback Propagation," in *ICLR 2025 Workshop XAI4Science: From Understanding Model Behavior to Discovering New Scientific Knowledge*, 2025.

[60] R. Achtibat, M. Dreyer, I. Eisenbraun, S. Bosse, T. Wiegand, W. Samek, and S. Lapuschkin, "From Attribution Maps to Human-Understandable Explanations through Concept Relevance Propagation," *Nature Machine Intelligence*, vol. 5, pp. 1006–1019, 2023.

[61] J.-H. Park, Y.-J. Ju, and S.-W. Lee, "Explaining generative diffusion models via visual analysis for interpretable decision-making process," *Expert Systems with Applications*, vol. 248, p. 123231, Aug. 2024.

[62] Y. Chen, L. Liu, and C. Ding, "X-iqe: explainable image quality evaluation for text-to-image generation with visual large language models," 2023.

[63] W. Baek, "attention-map-diffusers: Cross attention map tools for huggingface/diffusers." https://github.com/wooyeolbaek/attention-map-diffusers, 2024. MIT License.

[64] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," 2023.

[65] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer, F. Boesel, D. Podell, T. Dockhorn, Z. English, K. Lacey, A. Goodwin, Y. Marek, and R. Rombach, "Scaling rectified flow transformers for high-resolution image synthesis," *arXiv preprint arXiv:2403.03206*, 2024.

[66] B. F. Labs, "Flux." https://github.com/black-forest-labs/flux, 2024.

[67] Flower Federated Learning Framework, "Flower." https://flower.ai. Accessed: 2025-02-17.

[68] zhb2000, "Fedbox: Federated learning toolbox." https://github.com/zhb2000/fedbox. Accessed: 2025-02-17.

[69] C. Bucilu, R. Caruana, and A. Niculescu-Mizil, "Model compression," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 535–541, ACM, 2006.

[70] A. Bär, F. Hüger, P. Schlicht, and T. Fingscheidt, "On the robustness of redundant teacher-student frameworks for semantic segmentation," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1380–1388, 2019.

[71] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[72] J. Li, R. Zhao, J. T. Huang, and Y. Gong, "Learning small-size dnn with output-distribution-based criteria," in *Interspeech*, 2014.

[73] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," *CoRR*, vol. abs/1412.6550, 2014.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[74] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *ArXiv*, vol. abs/1612.03928, 2016.

[75] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," *ArXiv*, vol. abs/1610.02136, 2016.

[76] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[77] K. N. Mallikarjunan, K. Muthupriya, and S. M. Shalinie, "A survey of distributed denial of service attack," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1–6, IEEE, 2016.

[78] S. S. Shivakumar, N. Rodrigues, A. Zhou, I. D. Miller, V. Kumar, and C. J. Taylor, "Pst900: Rgb-thermal calibration, dataset and segmentation network," in *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 9441–9447, IEEE, 2020.

[79] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, "Aerial imagery pile burn detection using deep learning: The flame dataset," *Computer Networks*, vol. 193, p. 108001, 2021.

[80] D. Fotiou, V. Mygdalis, and I. Pitas, "Robofirefusenet: Robust fusion of visible and infrared wildfire imaging for real-time flame and smoke segmentation,"

[81] W. Zhou, J. Liu, J. Lei, L. Yu, and J.-N. Hwang, "Gmnet: Graded-feature multilabel-learning network for rgb-thermal urban scene semantic segmentation," *IEEE Transactions on Image Processing*, vol. 30, pp. 7790–7802, 2021.

[82] X. Ji, J. F. Henriques, and A. Vedaldi, "Invariant information clustering for unsupervised image classification and segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9865–9874, 2019.

[83] J. H. Cho, U. Mall, K. Bala, and B. Hariharan, "Picie: Unsupervised semantic segmentation using invariance and equivariance in clustering," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16794–16804, 2021.

[84] M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely, and W. T. Freeman, "Unsupervised semantic segmentation by distilling feature correspondences," in *International Conference on Learning Representations*.

[85] C. Kim, W. Han, D. Ju, and S. J. Hwang, "Eagle: Eigen aggregation learning for object-centric unsupervised semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3523–3533, 2024.

[86] H. S. Seong, W. Moon, S. Lee, and J.-P. Heo, "Leveraging hidden positives for unsupervised semantic segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19540–19549, 2023.

[87] M. D. Tzimas, V. Mygdalis, C. Papaioannidis, and I. Pitas, "Extreme weakly supervised binary semantic image segmentation via one-pixel supervision," Apr. 2025.

[88] I. Mademlis, A. Tefas, and I. Pitas, "A salient dictionary learning framework for activity video summarization via key-frame extraction," *Inf. Sci.*, vol. 432, pp. 319–331, 2018.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[89] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman, "Video summarization with long short-term memory," *ArXiv*, vol. abs/1605.08110, 2016.

[90] M. Rochan, L. Ye, and Y. Wang, "Video summarization using fully convolutional sequence networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 358–374, Springer, 2018.

[91] B. Gong, W.-L. Chao, K. Grauman, and F. Sha, "Diverse sequential subset selection for supervised video summarization," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, pp. 2069–2077, 2014.

[92] I. Kaseris, E. Apostolidis, V. Mezaris, and I. Patras, "Exploiting caption diversity for unsupervised video summarization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 1–5, 2022.

[93] M. Kaseris, I. Mademlis, and I. Pitas, "Adversarial unsupervised video summarization augmented with dictionary loss," in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 2683–2687, IEEE, 2021.

[94] B. Mahasseni, M. Lam, and S. Todorovic, "Unsupervised video summarization with adversarial lstm networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2982–2991, IEEE, 2017.

[95] J. Fajtl, H. S. Sokeh, V. Argyriou, D. Monekosso, and P. Remagnino, "Summarizing videos with attention," in *Asian Conference on Computer Vision (ACCV) Workshops*, pp. 39–54, Springer, 2018.

[96] W. Zhu, J. Lu, J. Li, and J. Zhou, "Dsnet: A flexible detect-to-summarize network for video summarization," *IEEE Transactions on Image Processing*, vol. 30, pp. 948–962, 2020.

[97] E. Apostolidis, G. Balaouras, V. Mezaris, and I. Patras, "Combining global and local attention with positional encoding for video summarization," in *2021 IEEE International Symposium on Multimedia (ISM)*, pp. 226–234, IEEE, 2021.

[98] E. Charalampakis, C. Papaioannidis, and I. Pitas, "Divide-and-summarize: Enhancing deep neural video summarization," May 2025.

[99] Z. Ji, K. Xiong, Y. Pang, and X. Li, "Video summarization with attention-based encoder-decoder networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 3035–3045, 2019.

[100] W. Zhu, J. Lu, Y. Han, and J. Zhou, "Learning multiscale hierarchical attention for video summarization," *Pattern Recognition*, vol. 122, p. 108312, 2022.

[101] G. Liang, Y. Lv, S. Li, S. Zhang, and Y. Zhang, "Video summarization with a convolutional attentive adversarial network," *Pattern Recognition*, vol. 131, p. 108840, 2022.

[102] Y. Zhang, Y. Liu, and R. Tao, "Vss-net: Visual semantic self-mining network for video summarization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[103] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[104] N. Nguyen and B. Quanz, "Temporal latent auto-encoder: A method for probabilistic multivariate time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 9117–9125, 2021.

[105] S. Du, T. Li, and S.-J. Horng, "Time series forecasting using sequence-to-sequence deep learning framework," *2018 9th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, pp. 171–176, 2018.

[106] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[107] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998–6008, 2017.

[108] T. Gangopadhyay, S. Y. Tan, Z. Jiang, R. Meng, and S. Sarkar, "Spatiotemporal attention for multivariate time series prediction and interpretation," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3560–3564, 2021.

[109] G. E. P. Box, G. M. Jenkins, and J. F. Macgregor, "Some recent advances in forecasting and control," *Journal of The Royal Statistical Society Series C-applied Statistics*, vol. 17, pp. 158–179, 1968.

[110] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," 2013.

[111] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *ArXiv*, vol. abs/1704.02971, 2017.

[112] L. Bai, L. Yao, C. Li, X. Wang, and C. Wang, "Adaptive graph convolutional recurrent network for traffic forecasting," *ArXiv*, vol. abs/2007.02842, 2020.

[113] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 95–104, ACM, 2018.

[114] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systemspart i: Deterministic non-linear systems," *International Journal of Control*, vol. 41, no. 2, pp. 303–328, 1985.

[115] G. Chatziparaskevas, I. Mademlis, and I. Pitas, "Generative representation learning in recurrent neural networks for causal timeseries forecasting," *IEEE Transactions on Artificial Intelligence*, vol. 5, pp. 6412–6425, 2024.

[116] Y. Liu, C. Gong, L. Yang, and Y. Chen, "Dstp-rnn: a dual-stage two-phase attention-based recurrent neural networks for long-term and multivariate time series prediction," *ArXiv*, vol. abs/1904.07464, 2019.

[117] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang, "Spectral temporal graph neural network for multivariate time-series forecasting," *ArXiv*, vol. abs/2103.07719, 2020.

[118] N. A. Muhadi, A. F. Abdullah, S. K. Bejo, M. R. Mahadi, and A. Mijić, "Deep learning semantic segmentation for water level estimation using surveillance camera," *Applied Sciences*, 2021.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[119] Y. Liang, X. Li, B. Tsai, Q. Chen, and N. H. Jafari, "V-floodnet: A video segmentation system for urban flood detection and quantification," *Environ. Model. Softw.*, vol. 160, p. 105586, 2022.

[120] P. Akiva, M. Purri, K. J. Dana, B. Tellman, and T. Anderson, "H2o-net: Self-supervised flood segmentation via adversarial domain adaptation and label refinement," *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 111–122, 2020.

[121] D. Hernández, J. M. Cecilia, J.-C. Cano, and C. M. T. Calafate, "Flood detection using real-time image segmentation from unmanned aerial vehicles on edge-computing platform," *Remote. Sens.*, vol. 14, p. 223, 2022.

[122] L. Yang, W. Zhuo, L. Qi, Y. Shi, and Y. Gao, "St++: Make self-trainingwork better for semi-supervised semantic segmentation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4258–4267, 2021.

[123] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, IEEE, 2017.

[124] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International Journal of Computer Vision*, vol. 129, pp. 3051 – 3068, 2020.

[125] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du, Q. Dang, B. Lai, Q. Liu, X. Hu, D. Yu, and Y. Ma, "Pp-liteseg: A superior real-time semantic segmentation model," *ArXiv*, vol. abs/2204.02681, 2022.

[126] Y. Hong, H. Pan, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes," *ArXiv*, vol. abs/2101.06085, 2021.

[127] N. Ma, J. Fan, W. Wang, J. Wu, Y. Jiang, L. Xie, and R. Fan, "Computer vision for road imaging and pothole detection: A state-of-the-art review of systems and algorithms," *ArXiv*, vol. abs/2204.13590, 2022.

[128] X. Liang, X. Yu, C. Chen, Y. Jin, and J. Huang, "Automatic classification of pavement distress using 3d ground-penetrating radar and deep convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 22269–22277, 2022.

[129] A. Ahmed, M. Ashfaque, M. U. UlHaq, S. Mathavan, K. Kamal, and M. Rahman, "Pothole 3d reconstruction with a novel imaging system and structure from motion techniques," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 4685–4694, 2021.

[130] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Transactions on Image Processing*, vol. 29, pp. 897–908, 2019.

[131] R. Fan, H. Wang, Y. Wang, M. Liu, and I. Pitas, "Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms," *IEEE Transactions on Image Processing*, vol. 30, pp. 8144–8154, 2021.

[132] S. Roy, "Stereo without epipolar lines: A maximum-flow formulation," *International Journal of Computer Vision*, vol. 34, pp. 147–161, 1999.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[133] C.-W. Liu, H. Wang, S. Guo, M. J. Bocus, Q. Chen, and R. Fan, "Stereo matching: Fundamentals, state-of-the-art, and existing challenges," in *Autonomous Driving Perception*, pp. 63–100, Springer Nature Singapore, 2023.

[134] J.-R. Chang and Y. Chen, "Pyramid stereo matching network," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5410–5418, 2018.

[135] U. Efe, K. G. Ince, and A. A. Alatan, "Dfm: A performance baseline for deep feature matching," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4279–4288, 2021.

[136] R. Fan, U. Ozgunalp, Y. Wang, M. Liu, and I. Pitas, "Rethinking road surface 3-d reconstruction and pothole detection: From perspective transformation to disparity map segmentation," *IEEE Transactions on Cybernetics*, vol. 52, pp. 5799–5808, 2020.

[137] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, pp. 211 – 252, 2014.

[138] A. Dosovitskiy, G. Ros, F. Codevilla, A. M. López, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on Robot Learning*, 2017.

[139] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3d reconstruction based on dense subpixel disparity map estimation," *IEEE Transactions on Image Processing*, vol. 27, pp. 3025–3035, 2018.

[140] R. Fan, J. Jiao, J. Pan, H. Huang, S. Shen, and M. Liu, "Real-time dense stereo embedded in a uav for road inspection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 535–543, 2019.

[141] H. Xu and J. Zhang, "Aanet: Adaptive aggregation network for efficient stereo matching," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1956–1965, 2020.

[142] B. Xu, Y. Xu, X. Yang, W. Jia, and Y. Guo, "Bilateral grid learning for stereo matching networks," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12492–12501, 2021.

[143] B. Liu, H. Yu, and Y. Long, "Local similarity pattern and cost self-reassembling for deep stereo matching networks," in *AAAI Conference on Artificial Intelligence*, 2021.

[144] H. Xu, J. Zhang, J. Cai, H. Rezatofighi, F. Yu, D. Tao, and A. Geiger, "Unifying flow, stereo and depth estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 13941–13958, 2022.

[145] J. Li, P. Wang, P. Xiong, T. Cai, Z.-P. Yan, L. Yang, J. Liu, H. Fan, and S. Liu, "Practical stereo matching via cascaded recurrent network with adaptive correlation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16242–16251, 2022.

[146] B. Liu, H. Yu, and G. Qi, "Graftnet: Towards domain generalized stereo matching with a broad-spectrum and task-oriented feature," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13002–13011, 2022.

[147] T. Chang, X. Yang, T. Zhang, M. Wang, and M. Y, "Domain generalized stereo matching via hierarchical visual transformation," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9559–9568, 2023.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[148] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "Yolov6: A single-stage object detection framework for industrial applications," *ArXiv*, vol. abs/2209.02976, 2022.

[149] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *ArXiv*, vol. abs/2004.10934, 2020.

[150] G. Jocher, A. Chaurasia, J. Qiu, and A. Hogan, "Ultralytics yolov8," 2023.

[151] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *ArXiv*, vol. abs/2107.08430, 2021.

[152] G. Jocher, "Yolov5 by ultralytics," 2020.

[153] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7464–7475, 2022.

[154] M. Chamikara, P. Bertok, I. Khalil, D. Liu, and S. Camtepe, "Privacy preserving face recognition utilizing differential privacy," *Computers & Security*, vol. 97, p. 101951, 2020.

[155] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements. arxiv 2024," *arXiv preprint arXiv:2410.17725*.

[156] Z. Chen, S. Li, B. Yang, Q. Li, and H. Liu, "Multi-scale spatial temporal graph convolutional network for skeleton-based action recognition," in *AAAI Conference on Artificial Intelligence*, 2021.

[157] Y. Chen, Z. Zhang, C. Yuan, B. Li, Y. Deng, and W. Hu, "Channel-wise topology refinement graph convolution for skeleton-based action recognition," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13339–13348, 2021.

[158] D. Chen, Y. Lin, W. Li, P. Li, J. Zhou, and X. Sun, "Measuring and relieving the over-smoothing problem for graph neural networks from the topological view," in *AAAI Conference on Artificial Intelligence*, 2019.

[159] A. F. Bavil, H. Damirchi, and H. D. Taghirad, "Action capsules: Human skeleton action recognition," *Comput. Vis. Image Underst.*, vol. 233, p. 103722, 2023.

[160] Y. Pang, Q. Ke, H. Rahmani, J. Bailey, and J. Liu, "Igformer: Interaction graph transformer for skeleton-based human interaction recognition," in *European Conference on Computer Vision*, 2022.

[161] A. M. D. Boissiere and R. Noumeir, "Infrared and 3d skeleton feature fusion for rgb-d action recognition," *IEEE Access*, vol. 8, pp. 168297–168308, 2020.

[162] A. Kamel, C. Zhang, and I. Pitas, "Spatio-temporal invariant descriptors for skeleton-based human action recognition," *Inf. Sci.*, vol. 700, p. 121832, 2025.

[163] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *AAAI Conference on Artificial Intelligence*, 2018.

[164] C. Si, Y. Jing, W. Wang, L. Wang, and T. Tan, "Skeleton-based action recognition with spatial reasoning and temporal stack learning," in *European Conference on Computer Vision*, 2018.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[165] M. Li, S. Chen, X. Chen, Y. Zhang, Y. Wang, and Q. Tian, "Actional-structural graph convolutional networks for skeleton-based action recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3590–3598, 2019.

[166] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12018–12027, 2018.

[167] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional lstm network for skeleton-based action recognition," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1227–1236, 2019.

[168] Y. Song, Z. Zhang, C. Shan, and L. Wang, "Richly activated graph convolutional network for robust skeleton-based action recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, pp. 1915–1925, 2020.

[169] K. Cheng, Y. Zhang, X. He, W. Chen, J. Cheng, and H. Lu, "Skeleton-based action recognition with shift graph convolutional network," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 180–189, 2020.

[170] F. Ye, S. Pu, Q. Zhong, C. Li, D. Xie, and H. Tang, "Dynamic gcn: Context-enriched topology learning for skeleton-based action recognition," *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.

[171] C. Plizzari, M. Cannici, and M. Matteucci, "Skeleton-based action recognition via spatial and temporal transformer networks," *Comput. Vis. Image Underst.*, vol. 208-209, p. 103219, 2020.

[172] K. Xu, F. Ye, Q. Zhong, and D. Xie, "Topology-aware convolutional neural network for efficient skeleton-based action recognition," *ArXiv*, vol. abs/2112.04178, 2021.

[173] Y. Song, Z. Zhang, C. Shan, and L. Wang, "Constructing stronger and faster baselines for skeleton-based action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 1474–1488, 2021.

[174] Y. Liu, H. Zhang, Y. Li, K. He, and D. Xu, "Skeleton-based human action recognition via large-kernel attention graph convolutional network," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, pp. 2575–2585, 2023.

[175] B. Nikpour and N. Armanfard, "Spatial hard attention modeling via deep reinforcement learning for skeleton-based human activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, pp. 4291–4301, 2023.

[176] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, Q. Jiang, C. Li, J. Yang, H. Su, J. Zhu, and L. Zhang, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," 2024.

[177] Y. Zhang, X. Li, H. Wang, and J. Chen, "Person re-identification network based on edge-enhanced feature extraction and inter-part relationship modeling," *Applied Sciences*, vol. 14, no. 18, p. 8244, 2024.

[178] S. Voigt, F. Giulio-Tonolo, J. Lyons, J. Kuera, B. Jones, T. Schneiderhan, G. Platzeck, K. Kaku, M. K. Hazarika, L. Czaran, S. Li, W. Pedersen, G. K. James, C. Proy, D. M. Muthike, J. Bequignon, and D. Guha-Sapir, "Global trends in satellite-based emergency mapping," *Science*, vol. 353, pp. 247–252, July 2016. Number: 6296.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[179] G. Mateo-Garcia, L. Gomez-Chova, and G. Camps-Valls, "Convolutional neural networks for multispectral image cloud masking," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (Fort Worth, TX), pp. 2255–2258, IEEE, 2017.

[180] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for Sentinel-1," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 835–845, June 2020. ISSN: 2160-7516.

[181] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina, "Deep learning methods for flood mapping: a review of existing applications and future research directions," *Hydrology and Earth System Sciences*, vol. 26, no. 16, pp. 4345–4378, 2022. Publisher: Copernicus GmbH.

[182] M. Wieland and S. Martinis, "A modular processing chain for automated flood monitoring from multi-spectral satellite data," *Remote Sensing*, vol. 11, no. 9, p. 2330, 2019. Number: 9.

[183] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[184] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137–1149, June 2017.

[185] M. Wieland, F. Fichtner, S. Martinis, S. Groth, C. Krullikowski, S. Plank, and M. Motagh, "S1S2-Water: A global dataset for semantic segmentation of water bodies from Sentinel-1 and Sentinel-2 data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 1084–1099, 2023.

[186] M. Wieland, S. Schmidt, B. Resch, A. Abecker, and S. Martinis, "Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods," *Natural Hazards*, Jan. 2025.

[187] D. Krause, E. Schwarz, S. Voinov, H. Damerow, and D. Tomecki, "Sentinel-1 near real-time application for maritime situational awareness," *CEAS Space Journal*, vol. 11, pp. 45–53, Mar. 2019.

[188] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric, Y. Bulatov, and B. McCord, "xView: Objects in Context in Overhead Imagery," p. 16.

[189] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, . Yifu), C. Wong, A. V, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022.

[190] E. Chuvieco, F. Mouillot, G. R. van der Werf, J. San Miguel, M. Tanase, N. Koutsias, M. García, M. Yebra, M. Padilla, I. Gitas, A. Heil, T. J. Hawbaker, and L. Giglio, "Historical background and current developments for mapping burned area from satellite Earth observation," *Remote Sensing of Environment*, vol. 225, pp. 45–64, 2019.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[191] S. Hantson, M. Padilla, D. Corti, and E. Chuvieco, "Strengths and weaknesses of modis hotspots to characterize global fire occurrence," *Remote sensing of Environment*, vol. 131, pp. 152–159, 2013.

[192] L. Boschetti, D. P. Roy, L. Giglio, H. Huang, M. Zubkova, and M. L. Humber, "Global validation of the collection 6 MODIS burned area product," *Remote Sensing of Environment*, vol. 235, p. 111490, 2019.

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

[194] S. Battiston, M. Friedemann, D. Gascón, A. Viseras, A. Cardil, M. A. I. Mendes, J. Vendrell, B. Barth, E. Nebot, S. Martinis, and S. Clandillon, "Heimdall: A technological solution for multi-hazard management support including wildfires," in *6th International Fire Behavior and Fuels Conference*, pp. 1–6, 2019.

[195] Y. Liu, W. E. Heilman, B. E. Potter, C. B. Clements, W. A. Jackson, N. H. F. French, S. L. Goodrick, A. K. Kochanski, N. K. Larkin, P. W. Lahm, T. J. Brown, J. P. Schwarz, S. M. Strachan, and F. Zhao, *Smoke Plume Dynamics*, pp. 83–119. Cham: Springer International Publishing, 2022.

[196] X. Wang, Y. Xu, S. Liu, B. Ren, J. Kosinka, A. C. Telea, J. Wang, C. Song, J. Chang, C. Li, J. J. Zhang, and X. Ban, "Physics-based fluid simulation in computer graphics: Survey, research trends, and challenges," *Computational Visual Media*, vol. 10, no. 5, pp. 803–858, 2024.

[197] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, "The FEniCS project version 1.5," *Archive of Numerical Software* 3, 2015.

[198] J. Forthofer and B. Butler, "A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. part i. model formulation and comparison against measurements," *International Journal of Wildland Fire*, vol. 23, pp. 969–981, 08 2014.

[199] R. Albtoush, R. Dobrescu, and F. Ionescu, "A hierarchical model for emergency management systems," *UPB Scientific Bulletin*, vol. 73, Jan. 2011.

[200] J. Phengsuwan, T. Shah, N. B. Thekkummal, Z. Wen, R. Sun, D. Pullarkatt, H. Thirugnanam, M. V. Ramesh, G. Morgan, P. James, and R. Ranjan, "Use of Social Media Data in Disaster Management: A Survey," *Future Internet*, vol. 13, p. 46, Feb. 2021.

[201] Z. Lei, Y. Dong, W. Li, R. Ding, Q. Wang, and J. Li, "Harnessing Large Language Models for Disaster Management: A Survey," Jan. 2025.

[202] E. Steiger, B. Resch, and A. Zipf, "Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks," *International Journal of Geographical Information Science*, vol. 30, pp. 1694–1716, Sept. 2016.

[203] B. Resch, F. Usländer, and C. Havas, "Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment," *Cartography and Geographic Information Science*, vol. 45, pp. 362–376, July 2018.

[204] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Mar. 2003.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[205] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, "Topic Modeling in Embedding Spaces," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 439–453, 2020.

[206] D. Angelov, "Top2Vec: Distributed Representations of Topics," Aug. 2020.

[207] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022.

[208] D. Hanny and B. Resch, "Clustering-Based Joint Topic-Sentiment Modeling of Social Media Data: A Neural Networks Approach," *Information*, vol. 15, p. 200, Apr. 2024.

[209] D. Hanny and B. Resch, "Multimodal Geo-Information Extraction from Social Media for Supporting Decision-Making in Disaster Management," *AGILE: GIScience Series*, vol. 5, pp. 1–8, May 2024.

[210] D. Hanny and B. Resch, "Multimodal geoai: An integrated spatio-temporal topic-sentiment model for the analysis of geo-social media posts for disaster management," *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, p. 104540, 2025.

[211] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, (Hong Kong, China), pp. 3980–3990, Association for Computational Linguistics, 2019.

[212] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, (Minneapolis, Minnesota), pp. 4171–4186, Association for Computational Linguistics, June 2019.

[213] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised Cross-lingual Representation Learning at Scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, eds.), (Online), pp. 8440–8451, Association for Computational Linguistics, July 2020.

[214] OpenAI, "Hello GPT-4o." https://openai.com/index/hello-gpt-4o/, May 2024.

[215] Y. Yang, D. Cer, A. Ahmad, M. Guo, J. Law, N. Constant, G. Hernandez Abrego, S. Yuan, C. Tar, Y.-h. Sung, B. Strope, and R. Kurzweil, "Multilingual Universal Sentence Encoder for Semantic Retrieval," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, (Online), pp. 87–94, Association for Computational Linguistics, July 2020.

[216] X. Wu, F. Pan, and A. T. Luu, "Towards the TopMost: A topic modeling system toolkit," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)* (Y. Cao, Y. Feng, and D. Xiong, eds.), (Bangkok, Thailand), pp. 31–41, Association for Computational Linguistics, Aug. 2024.

[217] S. Madichetty and M. Sridevi, "Detecting Informative Tweets during Disaster using Deep Neural Networks," in *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, (Bengaluru, India), pp. 709–713, IEEE, Jan. 2019.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[218] T. Papadimos, S. Andreadis, I. Gialampoukidis, S. Vrochidis, and I. Kompatsiaris, "Flood-Related Multimedia Benchmark Evaluation: Challenges, Results and a Novel GNN Approach," *Sensors*, vol. 23, p. 3767, Apr. 2023.

[219] S. Madichetty, S. Muthukumarasamy, and P. Jayadev, "Multi-modal classification of Twitter data during disasters for humanitarian response," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 10223–10237, Nov. 2021.

[220] A. Olteanu, S. Vieweg, and C. Castillo, "What to Expect When the Unexpected Happens: Social Media Communications Across Crises," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, (New York, NY, USA), pp. 994–1009, Association for Computing Machinery, Feb. 2015.

[221] E. Blomeier, S. Schmidt, and B. Resch, "Drowning in the Information Flood: Machine-Learning-Based Relevance Classification of Flood-Related Tweets for Disaster Management," *Information*, vol. 15, p. 149, Mar. 2024.

[222] X. Zhang, Y. Malkov, O. Florez, S. Park, B. McWilliams, J. Han, and A. El-Kishky, "TwHIN-BERT: A Socially-Enriched Pre-trained Language Model for Multilingual Tweet Representations at Twitter," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, (New York, NY, USA), pp. 5597–5607, Association for Computing Machinery, Aug. 2023.

[223] B. Parhami, "Voting algorithms," *IEEE Transactions on Reliability*, vol. 43, no. 4, pp. 617–629, 1994.

[224] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," *30th Annual Symposium on Foundations of Computer Science*, pp. 256–261, 1989.

[225] A. P. Dawid and A. Skene, "Maximum likelihood estimation of observer errorrates using the em algorithm," *Journal of The Royal Statistical Society Series C-applied Statistics*, vol. 28, pp. 20–28, 1979.

[226] D. Cai, D. C. Nguyen, S. H. Lim, and L. Wynter, "Variational bayesian inference for crowdsourcing predictions," *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 3166–3172, 2020.

[227] R. Snow, B. T. O'Connor, D. Jurafsky, and A. Ng, "Cheap and fast but is it good? evaluating non-expert annotations for natural language tasks," in *Conference on Empirical Methods in Natural Language Processing*, 2008.

[228] F. Rodrigues and F. C. Pereira, "Deep learning from crowds," in *AAAI Conference on Artificial Intelligence*, 2017.

[229] F. Avgoustidis, P. Bassia, and I. Pitas, "Trustworthy majority voting for labeling and analyzing multi-annotator text sentiment datasets," Apr. 2025.

[230] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[231] P. Ekman, "Are there basic emotions?," *Psychological review*, vol. 99 3, pp. 550–3, 1992.

[232] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," 2013.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[233] V. K. Jain, S. Kumar, and S. L. Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *Journal of Computational Science*, vol. 21, pp. 316–326, 2017.

[234] M. Hasan, E. Rundensteiner, and E. Agu, "Automatic emotion detection in text streams by analyzing twitter data," *International Journal of Data Science and Analytics*, vol. 7, pp. 35–51, Feb. 2019.

[235] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, pp. 309–317, 2019.

[236] Z. Ahmad, R. Jindal, A. Ekbal, and P. Bhattachharyya, "Borrow from rich cousin: Transfer learning for emotion detection using cross lingual embedding," *Expert Systems with Applications*, vol. 139, p. 112851, 2020.

[237] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, 2021.

[238] H. Luo, L. Ji, T. Li, D. Jiang, and N. Duan, "Grace: Gradient harmonized and cascaded labeling for aspect-based sentiment analysis," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, (Online), pp. 54–64, Association for Computational Linguistics, 2020.

[239] C. Zorenböhmer, S. Schmidt, and B. Resch, "Emograce: Aspect-based emotion analysis for social media data," 2025.

[240] C. Zorenböhmer, S. Gandhi, S. Schmidt, and B. Resch, "An aspect-based emotion analysis approach on wildfire-related geo-social media data: A case study of the 2020 california wildfires," in *GWF 2025 ESG Summit: Geospatial Innovation for Risk Management, Business Transformation and Sustainability*, 2025.

[241] A. Getis and J. K. Ord, "The analysis of spatial association by use of distance statistics," *Geographical Analysis*, vol. 24, no. 3, pp. 189–206, 1992.

[242] L. Anselin, "Local indicators of spatial association—lisa," *Geographical Analysis*, vol. 27, no. 2, pp. 93–115, 1995.

[243] X. Zhang, L. Liu, L. Xiao, and J. Ji, "Comparison of machine learning algorithms for predicting crime hotspots," *IEEE Access*, vol. 8, pp. 181302–181310, 2020.

[244] S. Schmidt, M. Friedemann, D. Hanny, B. Resch, T. Riedlinger, and M. Mühlbauer, "Enhancing satellite-based emergency mapping: Identifying wildfires through geo-social media analysis," *Big Earth Data*, pp. 1–23, 2025.

[245] B. Barz, K. Schröter, A.-C. Kra, and J. Denzler, "Finding relevant flood images on twitter using content-based filters," *ICPR International Workshops and Challenges. ICPR 2021. Lecture Notes in Computer Science*, vol. 12666, 2021.

[246] S. Z. Hassan, K. Ahmad, S. Hicks, P. Halvorsen, A. Al-Fuqaha, N. Conci, and M. Riegler, "Visual sentiment analysis from disaster images in social media," *Sensors*, vol. 22, no. 10, p. 3628, 2022.

[247] Y. Li and Y. Xie, "Is a picture worth a thousand words? an empirical study of image content and social media engagement," *Journal of Marketing Research*, vol. 57, no. 1, pp. 1–19, 2020.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

[248] D. Zhu, J. Chen, K. Haydarov, X. Shen, W. Zhang, and M. Elhoseiny, "Chatgpt asks, blip-2 answers: Automatic questioning towards enriched visual descriptions," 2023.

[249] Y. Liu, Z. Li, M. Huang, B. Yang, W. Yu, C. Li, X. Yin, C.-l. Liu, L. Jin, and X. Bai, "Ocrbench: On the hidden mystery of ocr in large multimodal models," *Science China Information Sciences*, vol. 67, p. 220102, Dec. 2024.

[250] A. Chen, Z. Wang, C. Dong, K. Tian, R. Zhao, X. Liang, Z. Kang, and X. Li, "Chinaopen: A dataset for open-world multimodal learning," 2023.

[251] L. Juhász, P. Mooney, H. H. Hochmair, and B. Guan, "Chatgpt as a mapping assistant: A novel method to enrich maps with generative ai and content derived from street-level photographs," 2023.

[252] J. Li, D. Li, S. Savarese, and S. Hoi, "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *Proceedings of the 40th International Conference on Machine Learning* (A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, eds.), vol. 202 of *Proceedings of Machine Learning Research*, pp. 19730–19742, PMLR, 23–29 Jul 2023.

[253] A. Awadalla *et al.*, "Openflamingo: An open-source framework for training large autoregressive vision-language models," 2023.

[254] H. Liu *et al.*, "Visual instruction tuning," in *Advances in Neural Information Processing Systems*, 2023.

[255] B. Xiao, H. Wu, W. Xu, X. Dai, H. Hu, Y. Lu, M. Zeng, C. Liu, and L. Yuan, "Florence-2: Advancing a unified representation for a variety of vision tasks," *arXiv preprint arXiv:2311.06242*, 2023.

[256] S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh, "Vqa: Visual question answering," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2425–2433, 2015.

[257] M. Abdin, J. Aneja, H. Awadalla, A. Awadallah, A. A. Awan, N. Bach, A. Bahree, A. Bakhtiari, J. Bao, H. Behl, A. Benhaim, M. Bilenko, J. Bjorck, S. Bubeck, M. Cai, Q. Cai, V. Chaudhary, D. Chen, D. Chen, W. Chen, Y.-C. Chen, Y.-L. Chen, H. Cheng, P. Chopra, X. Dai, M. Dixon, R. Eldan, V. Fragoso, J. Gao, M. Gao, M. Gao, A. Garg, A. D. Giorno, A. Goswami, S. Gunasekar, E. Haider, J. Hao, R. J. Hewett, W. Hu, J. Huynh, D. Iter, S. A. Jacobs, M. Javaheripi, X. Jin, N. Karampatziakis, P. Kauffmann, M. Khademi, D. Kim, Y. J. Kim, L. Kurilenko, J. R. Lee, Y. T. Lee, Y. Li, Y. Li, C. Liang, L. Liden, X. Lin, Z. Lin, C. Liu, L. Liu, M. Liu, W. Liu, X. Liu, C. Luo, P. Madan, A. Mahmoudzadeh, D. Majercak, M. Mazzola, C. C. T. Mendes, A. Mitra, H. Modi, A. Nguyen, B. Norick, B. Patra, D. Perez-Becker, T. Portet, R. Pryzant, H. Qin, M. Radmilac, L. Ren, G. de Rosa, C. Rosset, S. Roy, O. Ruwase, O. Saarikivi, A. Saied, A. Salim, M. Santacroce, S. Shah, N. Shang, H. Sharma, Y. Shen, S. Shukla, X. Song, M. Tanaka, A. Tupini, P. Vaddamanu, C. Wang, G. Wang, L. Wang, S. Wang, X. Wang, Y. Wang, R. Ward, W. Wen, P. Witte, H. Wu, X. Wu, M. Wyatt, B. Xiao, C. Xu, J. Xu, W. Xu, J. Xue, S. Yadav, F. Yang, J. Yang, Y. Yang, Z. Yang, D. Yu, L. Yuan, C. Zhang, C. Zhang, J. Zhang, L. L. Zhang, Y. Zhang, Y. Zhang, Y. Zhang, and X. Zhou, "Phi-3 technical report: A highly capable language model locally on your phone," 2024.

[258] M. Deitke, C. Clark, S. Lee, R. Tripathi, Y. Yang, J. S. Park, M. Salehi, N. Muennighoff, K. Lo, L. Soldaini, J. Lu, T. Anderson, E. Bransom, K. Ehsani, H. Ngo, Y. Chen, A. Patel, M. Yatskar, C. Callison-Burch, A. Head, R. Hendrix, F. Bastani, E. VanderBilt, N. Lambert,

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Y. Chou, A. Chheda, J. Sparks, S. Skjonsberg, M. Schmitz, A. Sarnat, B. Bischoff, P. Walsh, C. Newell, P. Wolters, T. Gupta, K.-H. Zeng, J. Borchardt, D. Groeneveld, C. Nam, S. Lebrecht, C. Wittlif, C. Schoenick, O. Michel, R. Krishna, L. Weihs, N. A. Smith, H. Hajishirzi, R. Girshick, A. Farhadi, and A. Kembhavi, "Molmo and pixmo: Open weights and open data for state-of-the-art vision-language models," 2024.

[259] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, K. Dang, P. Wang, S. Wang, J. Tang, H. Zhong, Y. Zhu, M. Yang, Z. Li, J. Wan, P. Wang, W. Ding, Z. Fu, Y. Xu, J. Ye, X. Zhang, T. Xie, Z. Cheng, H. Zhang, Z. Yang, H. Xu, and J. Lin, "Qwen2.5-vl technical report," 2025.

[260] S. Schmidt, E. Díaz Fragachan, D. Arifi, D. Hanny, and B. Resch, "Assessing the spatial accuracy of geocoding flood-related imagery using vision language models," *Spatial Information Research*, vol. 33, no. 2, p. 15, 2025.

[261] F. Barbieri, L. Espinosa Anke, and J. Camacho-Collados, "XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, (Marseille, France), pp. 258–266, European Language Resources Association, June 2022.

[262] A. Olteanu, C. Castillo, F. Diaz, and S. Vieweg, "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, pp. 376–385, May 2014.

[263] D. Hanny, S. Schmidt, and B. Resch, "Active Learning for Identifying Disaster-Related Tweets: A Comparison with Keyword Filtering and Generic Fine-Tuning," in *Intelligent Systems and Applications* (K. Arai, ed.), (Cham), pp. 126–142, Springer Nature Switzerland, 2024.

[264] M. Wieland, V. Hertel, C. Geiss, S. Martinis, and K. Lechner, "Toward Scalable Damage Assessment for Rapid Disaster Response," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3902–3905, July 2024. ISSN: 2153-7003.

[265] C. Gilga, C. Hochwarter, L. Knoche, S. Schmidt, G. Ringler, M. Wieland, B. Resch, and B. Wagner, "Legal and ethical considerations for demand-driven data collection and AI-based analysis in flood response," *International Journal of Disaster Risk Reduction*, vol. 122, p. 105441, May 2025.

[266] S. Dujardin, D. Arifi, S. Schmidt, C. Linard, and B. Resch, "Tracing online flood conversations across borders: A watershed level analysis of geo-social media topics during the 2021 european flood," 2024.

This project has received funding from the European Union – European Commission under grant agreement No 101093003

Annex A: Related publications and technical reports

No.	Title	Reference	Type
1	Multimodal Geo-Information Extraction from Social Media for Supporting Decision-Making in Disaster Management	[209]	Article
2	Clustering-Based Joint Topic-Sentiment Modeling of Social Media Data: A Neural Networks Approach	[208]	Article
3	Enhancing satellite-based emergency mapping: Identifying wildfires through geo-social media analysis	[244]	Article
4	Assessing the spatial accuracy of geocoding flood-related imagery using Vision Language Models	[260]	Article
5	Active Learning for Identifying Disaster-Related Tweets: A Comparison with Keyword Filtering and Generic Fine-Tuning	[263]	Article
6	EmoGRACE: Aspect-Based Emotion Analysis for Social Media Data	[239]	Preprint
7	Trustworthy Majority Voting for Labeling and Analyzing Multi-Annotator Text Sentiment Datasets	[229]	Article
8	Toward Scalable Damage Assessment for Rapid Disaster Response	[264]	Article
9	Fusion of geospatial information from remote sensing and social media to prioritise rapid response actions in case of floods	[186]	Article
10	Legal and ethical considerations for demand-driven data collection and AI-based analysis in flood response	[265]	Article
11	A Decentralized Sharding BFT Consensus Approach, for Efficient Decentralized DNN Inference	[23]	Article
12	Real-Time Flood Water Segmentation with Deep Neural Networks	[25]	Article
13	These Maps Are Made by Propagation: Adapting Deep Stereo Networks to Road Scenarios With Decisive Disparity Diffusion	[7]	Article
14	Navigating Neural Space: Revisiting Concept Activation Vectors to Overcome Directional Divergence	[48]	Article
15	Ensuring Medical AI Safety: Explainable AI-Driven Detection and Mitigation of Spurious Model Behavior and Associated Data	[50]	Preprint
16	Beyond Scalars: Concept-Based Alignment Analysis in Vision Transformers	[51]	Article
17	Post-Hoc Concept Disentanglement: From Correlated to Isolated Concept Representations	[52]	Article
18	CoSy: Evaluating Textual Explanations of Neurons	[53]	Article

This project has received funding from the European Union – European Commission under grant agreement No 101093003

19	FADE: Why Bad Descriptions Happen to Good Features	[54]	Preprint
20	Efficient and Flexible Neural Network Training through Layer-wise Feedback Propagation	[59]	Article
21	Quanda: An Interpretability Toolkit for Training Data Attribution Evaluation and Beyond	[55]	Article
22	A Close Look at Decomposition-based XAI-Methods for Transformer Language Models	[56]	Preprint
23	The Meta-Evaluation Problem in Explainable AI: Identifying Reliable Estimators with MetaQuantus	[32]	Article
24	A Fresh Look at Sanity Checks for Saliency Maps	[33]	Article
25	Reactive Model Correction: Mitigating Harm to Task-Relevant Features via Conditional Bias Suppression	[36]	Article
26	From Hope to Safety: Unlearning Biases of Deep Models via Gradient Penalization in Latent Space	[37]	Article
27	PURE: Turning Polysemantic Neurons Into Pure Features by Identifying Relevant Circuits	[41]	Article
28	Understanding the (Extra-)Ordinary: Validating Deep Model Decisions with Prototypical Concept-based Explanations	[38]	Article
29	Explainable AI for Time Series via Virtual Inspection Layers	[43]	Article
30	Explaining predictive uncertainty by exposing second-order effects	[42]	Article
31	DualView: Data Attribution from the Dual Perspective	[47]	Preprint
32	Tracing online flood conversations across borders: A watershed level analysis of geo-social media topics during the 2021 European flood	[266]	Preprint
33	Multimodal GeoAI: An integrated spatio-temporal topic-sentiment model for the analysis of geo-social media posts for disaster management	[210]	Article
34	RoboFireFuseNet: Robust Fusion of Visible and Infrared Wildfire Imaging for Real-Time Flame and Smoke Segmentation	[80]	Preprint
35	Extreme Weakly Supervised Binary Semantic Image Segmentation via One-Pixel Supervision	[87]	Preprint
36	Divide-and-Summarize: Enhancing Deep Neural Video Summarization	[98]	Article
37	Generative representation learning in recurrent neural networks for causal timeseries forecasting	[115]	Article
38	Improve Real-Time Flood Segmentation by Encoding and Distilling Foreground Information	[26]	Article
39	A Weighting Loss Approach for Transformer-Based Object Detection	[27]	Article
40	Spatio-temporal invariant descriptors for skeleton-based human action recognition	[162]	Article

This project has received funding from the European Union – European Commission under grant agreement No 101093003