Reusable, Sharable, and Transferable Smart Data Platform for Collaborative Development of Data-Driven Smart Cities

Background :

- UN's Sustainable Development Goals (https://sdgs.un.org/goals)
  - Goal 11: Make cities and human settlements inclusive, safe, resilient and sustainable
  - Goal 13: Take urgent action to combat climate change and its impacts
  - Goal 17: Strengthen the means of implementation and revitalize the global partnership for sustainable development
- Society 5.0 (https://www8.cao.go.jp/cstp/english/society5_0/index.html)
  - "A human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space."

Targets:

- Disaster risk reduction and sustainable transport
  - Smart environmental Tourism and Sustainable Mobility (Dalat city, Vietnam)
  - Smart Dengue Early Warning System (Cauayan City, Philippines)
  - Smart Outdoor Activities (Singapore)
- Atmosphere and Climate Change
  - Transboundary Air-Pollution Forecasting (Brunei)
- Multi-stakeholder partnerships and voluntary commitments
  - xData Platform and Event Data Sharing: A decentralized and collaborative approach to fast, economically, and sustainably develop user-centered applications.

Speaker: Minh-Son Dao

Reusable, Sharable, and Transferable Smart Data Platform for Collaborative Development of Data-Driven Smart Cities

Project Members:

Principal Investigator

- Senior.Researcher.Dr. **Minh-Son Dao** (National Institute of Information and Communications Technology, Japan)

Work-package Leaders

- Asst.Prof.Dr. **Asem Kasem** (Universiti Teknologi Brunei, Brunei)
- Lecturer.Dr. **Thanh-Hai Dang** (Dalat University, Vietnam)
- Asst.Prof.Dr. **Filip Biljecki** (National University of Singapore, Singapore)
- Prof. Betchie **Aguinaldo** (Isabela State University Cauayan Campus, Philippines)

Official Members

- Senior.Researcher.Dr. **Sadanori Ito** (National Institute of Information and Communications Technology, Japan)
- Lecturer.MSc. **Minh-Hiep Nguyen** (Dalat University, Vietnam)
- Lecturer.Dr. **Phuong-Binh Vo** (Dalat University, Vietnam)
- Asst.Prof.Dr. **Joel M. Gumiran** (Isabela State University Cauayan Campus, Philippines)
- Asst.Prof.Dr. **Cherry R Gumiran** (Isabela State University Cauayan Campus, Philippines)
- Assoc.Prof.Dr**. Arnel C Fajardo** (Isabela State University Cauayan Campus, Philippines)

Project Duration:

- 24 months (2020/04/01 – 2022/03/31)

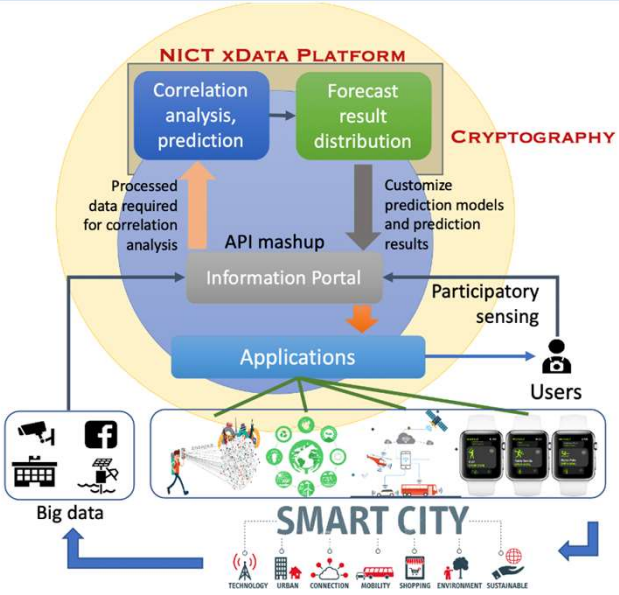
Project Budget:

- 80,000 USD

Project Activities: Scientific and Technological Development

1. Scientific

- Values and Complex Events Prediction
- Insights from cross-data
- Decentralized and collaborative development
- Data Visualization and Geo-based Navigation



2. Technological development

- Deep Multimodal Learning for Predicting Haze Transboundary using peripheral and weather from stations and open data.
- Discovering Periodic-Frequent Patterns in Very Large Uncertain Temporal Databases.
- Cross-Data model, Fuzzy membership and Visual pollution Integrating for Air pollution Estimation and Prediction using Images.
- Interactive Incident Retrieval System for First-view Travel-log Data with new incident classes automatically generation engine.
- Graph-based CCTV network dataset and baseline methods for Traffic Flow.
- Data collection system (e.g., sensor networks, crowdsourcing).
- A risk-avoidable navigation system based on cross-data.
- A Federated Learning Framework for Enabling Contextual Health Monitoring in a Cloud-Edge Network

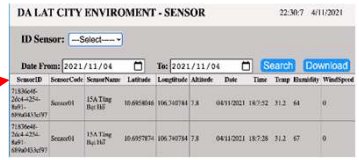
Project Activities: Experiments including field testing

- Sensor Networks designation and deployment
  - Environment-mobility sensor networks in Dalat city (Vietnam)
  - Dengue sensor networks in Cauayan city (Philippines)
- Data structure and database designation based on 3D-GIS data format to harmonize data exchanged among partners and to further adapt to smart cities (Singapore)
- Data collection:
  - Collect haze trans-border-related datasets (air pollution, fire forest, weather) from ASEAN countries (UTB)
  - Collect Dengue-related dataset in Cauayan city, the Philippines (ISU)
  - Collect CCTV videos and environment-weather dataset in Dalat city, Vietnam (DLU)
  - Collect first-view travel-log data for 8 typical road incidents from open sources (NICT)
- Transfer knowledge:
  - Transfer models and tools developed by NICT to partners for reusing on local data.
- Models Evaluation:
  - Evaluate the adaptability and accuracy of transferred models on local data (UTB, NICT, DLU, NUS)
- System Manual Instruction:
  - Construct the website to give manual instructions for reusing and sharing products developed by partners (https://www.xdata.nict.jp/xDataPFDocs/WebAPI/1.0/en/)

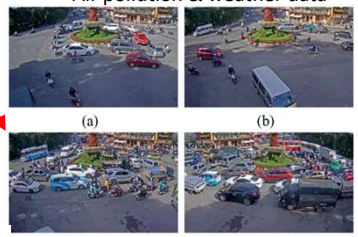
R&D results: Application (or system) development

Smart environmental-mobility (DLU)

Map-based Data Visualization



Air pollution & weather data



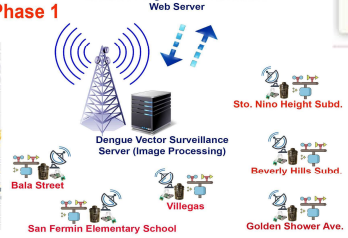
Video data

Urban Air Mobility Benchmark

Smart sustainable city

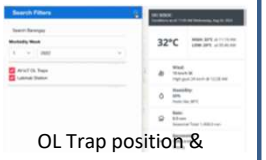
Risk map (app)

Community Dengue Early Warning System (ISU)



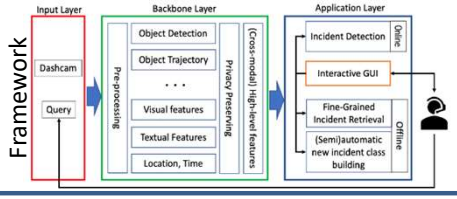
Dengue Vector Surveillance Website

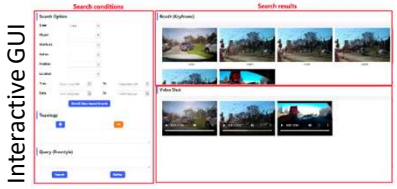
IoT OL Trap



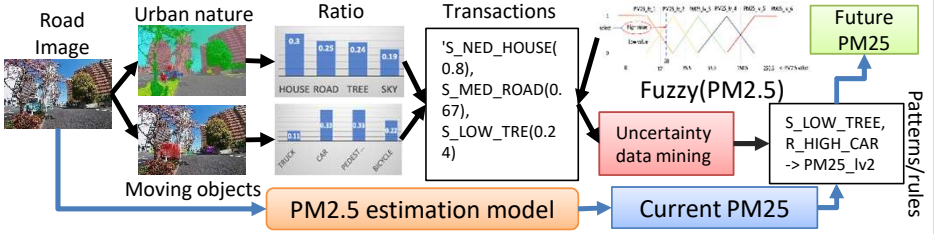
OL Trap position & weather data

Incident detection and retrieval from dashcams DB (NICT)





Anticipate Future AQI based on visual pollution (NICT)

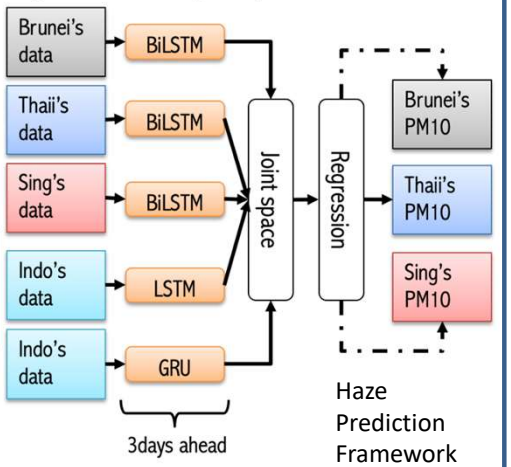


Haze transboundary Prediction (UTB)

Haze Transboundary System

Public DB

Baseline Algos



Insights for wellbeing challenge in Haze Transboundary

Benchmark

R&D results: Experiments including field testing

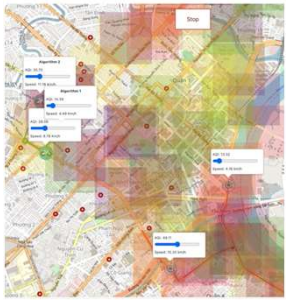
Haze transboundary (UTB)

Smart environmental-mobility (DLU)

Congestion Prediction using CCTV data

Model	Sensor Raw		Sensor + TL		Sensor + TLPW	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
CatBoost	61.09	59.33	70.82	71.20	86.93	87.35
Extra Trees	63.22	62.24	69.60	69.66	84.50	84.93
Linear Regression	42.25	32.55	42.55	34.55	49.54	47.47
Random Forest	65.65	65.40	71.43	71.43	86.32	86.64
SVM	59.27	58.44	63.83	65.09	80.24	81.47
XGBoost	**70.21**	**70.53**	**77.20**	**77.40**	87.54	87.70
Stacking	62.31	60.98	72.64	72.85	**87.84**	**88.13**

Air Pollution Prediction and Route Recommendation using **Urban Air Mobility** data

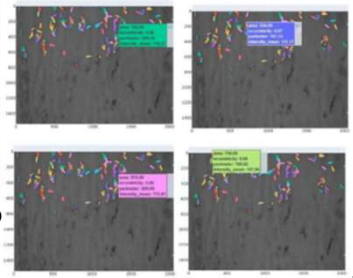


• image log

• Personal Environmental Sensors (PM2.5, PM10, O3, NO2, SO2, CO, UV, humidity, temperature, GPS)

Community Dengue Early Warning System (ISU)

IoT OL Trap Product



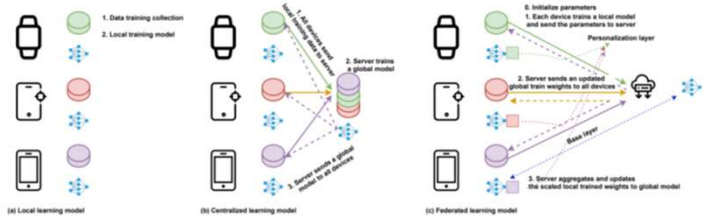
Detected Aedes aegypti egg with their morphological properties

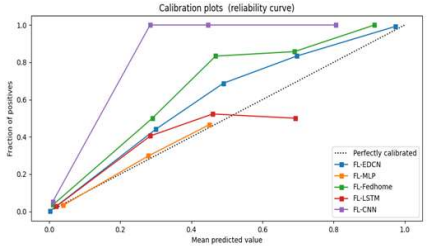
Weather and IoT OL Trap installation

	area	eccentricity	perimeter	centroid-0	centroid-1	orientation	axis_major_length	axis_minor_length
0	799	0.971879	213.515765	33.983730	43.167710	-0.091697	70.868184	16.688103
1	728	0.905050	246.829473	29.681319	310.265110	0.327946	56.150319	23.881044
2	479	0.983010	149.367532	27.421712	496.175365	0.425557	60.751977	11.151155
3	233	0.970838	101.426407	13.939914	520.446352	0.108693	40.694948	9.756136
4	668	0.734364	289.906638	25.389222	595.767964	0.896714	60.119735	40.806647

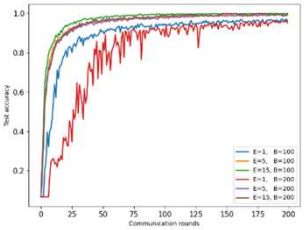
Number of Aedes aegypti egg detected with their morphological properties.

A Federated Learning Framework in a Cloud-Edge Network (NICT)





Comparison of calibration of FL model's partition



Comparison of accuracy vs. comm. rounds with different batch & epochs

Incident detection and retrieval from dashcams DB (NICT)

TABLE III
RESULTS ON VIDEO

Incident	F1-score
Animals	0.9031
Collapse	0.7811
Crash	0.7620
Fire	1.0000
Flooding	1.0000
Landslide	0.7508
Snow	0.8749
Treefall	0.7877

Table 3. Conditional-LSTM-NN model - 7 days look back - Input: public sensing data + personal images - Ground truth: sensor box (PM2.5) - Predict PM2.5

Input	Constraints	MAE Tokyo[12]	MAE MNR-HCM[13]	MAE VisionAir[14]
Image PM25, Image, Weather PM25, Image PM25	Both	4.089 16.928 **4.064** 21.741	21.28 12.06 **9.23** 29.34	7.19 12.42 9.55 24.80
Image PM25, Image, Weather PM25, Image PM25	Time	4.111 17.664 4.107 22.485	19.51 14.21 10.81 28.90	9.81 19.39 7.11 28.42
Image PM25, Image, Weather PM25, Image PM25	Location	4.117 16.916 4.076 21.771	19.48 16.34 9.48 20.00	13.73 11.50 **5.82** 21.89
Image PM25, Image, Weather PM25, Image PM25	None	4.120 17.136 19.573 21.892	18.53 16.06 10.24 27.35	10.86 14.22 18.17 24.33
#images		119,928	6,566	6,057

Models	P@10	P@50	P@100	P@200
Dual Encoder	0.5	0.64	0.60	0.68
SCAN	0.7	0.80	0.67	0.71
DSRAN	0.8	0.90	0.77	0.81
VSRN	0.6	0.74	0.63	0.40
Proposed	0.7	**0.94**	0.70	**0.85**

Dataset: BDD1000K (12K samples) + RetroTruck (254 samples) + I4W (600 samples)

A broken-down tree on the side of a road

Text query

Top-10

Assessing bikeability with street view (NUS)

Target_variable	MAE	MAPE	RMSE	R2
beauty	0.63	0.10	0.81	-0.20
building_attractiveness	0.63	0.10	0.78	-0.14
cleanliness	0.63	0.10	0.79	-0.12
cycling_attractiveness	0.65	0.11	0.83	-0.18
living_attractiveness	0.69	0.12	0.87	-0.10
safety	0.72	0.12	0.89	-0.29
spaciousness	0.67	0.11	0.83	-0.19

Predictive modeling of perception indicator

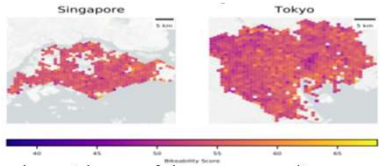


illustration of an urban setting together with one of the corresponding street-level views, highlighting several aspects that may indicate bikeability

R&D results: Scientific and technological

1. Values and Complex Events Prediction
   - MM-trafficEvent: Automatically create a new incident class, detect, and interactively retrieve incidents from dashcam videos.
   - MM-Mobility-AQI: Predict environmental quality using crossmodal, fuzzy, and uncertain temporal transaction data mining.
   - MM-trafficNET: Predict congestions using GNN techniques and CCTV videos.
   - MM-hazePrediction: Predict air pollution using Multi-Attention Spatio-Temporal Graph Networks
   - Fed xData: A Federated Learning Framework for Enabling Contextual Health Monitoring in a Cloud-Edge Network
   - Aedes Aegypti Egg Detection using Morphology and Computer Vision.
2. Insights from cross-data
   - Mining of Periodic-Frequent Patterns in Very Large Uncertain Temporal Databases, applied for Japan and Vietnam environment-visual dataset. The patterns can reason the association between urban nature, moving vehicles, and air pollution.
3. Decentralized and collaborative development
   - Transferred Learning: Utilize transfer learning models developed by NICT to partners on local datasets.
   - MM-sensing system: the integration system to connect personal multimedia devices, xData Edge (local servers) and xData PF for predicting events (e.g., AQI, congestion)
4. Data Visualization and Geo-based Navigation
   - A risk-avoidance navigation system based on the outputs mentioned models and algorithms for sustainable smart cities
   - Assessing bikeability with street view

Scientific Contribution: Presentations at International Conferences

No	Conference Paper Titles	Author names	Affiliation	Conference name	Date	Venue
1	Image-2-AQI: Aware of the Surrounding Air Qualification by a Few Images	**Minh-Son Dao***, K. Zettsu, Rage Uday Kiran	*NICT	IEA/AIE (rank B)	26-29/07/2021	Virtual Conference
2	MNR-Air: An Economic and Dynamic Crowdsourcing Mechanism to Collect Personal Lifelog and Surrounding Environment Dataset. A Case Study in Ho Chi Minh City, Vietnam	D.H. Nguyen, T.L.. Nguyen-Tai, M.T. Nguyen, T.B. Nguyen, **Minh-Son Dao***	*NICT	MMM (rank B)	22-24/01/2021	Virtual Conference
3	An Effective AQI Estimation Using Sensor Data and Stacking Mechanism	Q.D. Duong, M.Q. Le, T.L. Nguyen-Tai, D.H. Nguyen, **Minh-Son Dao***, T.B. Nguyen	*NICT	SoMET	21-23/09/2021	Virtual Conference
4	Discovering Spatial High Utility Itemsets in High-Dimensional Spatiotemporal Databases	S.C. Bommisetty, P. Ravikumar, R. Uday Kiran, **Minh-Son Dao***, Koji Zettsu	*NICT	IEA/AIE (rank B)	26-29/07/2021	Virtual Conference
5	Efficient Discovery of Partial Periodic-Frequent Patterns in Temporal Databases	S. Nakamura, R. Uday Kiran, L. Palla, P. Ravikumar, Y. Watanobe, **Minh-Son Dao***, K. Zettsu, M. Toyoda	*NICT	DEXA (rank B)	27-30/09/2021	Virtual Conference
6	Overview of MediaEval 2021: Insights for Wellbeing Task : Cross-Data Analytics for Transboundary Haze Prediction	**Asem Kasem*, Minh-Son Dao**, Effa Nabilla Aziz***, D.T. Dang-Nguyen, C. Gurrin, M.T. Tran, T.B. Nguyen, **Wida Suhaili***	*UTB, **NICT	MediaEval	12-14/12/2021	Virtual Conference
7	Discovering Periodic-Frequent Patterns in Very Large Uncertain Temporal Databases	L. Palla, R. Uday Kiran, **Minh-Son Dao***	*NICT	ICONIP (rank A)	08-12/12/2021	Virtual Conference
8	MM-trafficEvent: An Interactive Incident Retrieval System for First-view Travel-log Data	**Minh-Son Dao*,** Phu Nguyen, Duy Pham, Binh Nguyen, Koji Zettsu	*NICT	IEEE Big Data (rank B)	15-18/12/2021	Virtual Conference
9.	Improving the awareness of sustainable smart cities by analyzing lifelog images and IoT Pollution Data	Tuan-Vinh La, **Minh-Son Dao***, Kazuki Tejima, Rage Uday Kiran, Koji Zettsu	*NICT	IEEE Big Data (rank B)	15-18/12/2021	Virtual Conference
10.	Fed xData: A Federated Learning Framework for Enabling Contextual Health Monitoring in a Cloud-Edge Network	Tran Anh Khoa, Do-Van Nguyen, **Minh-Son Dao***, Koji Zettsu	*NICT	IEEE Big Data (rank B)	15-18/12/2021	Virtual Conference
11	A Robust Ensemble Method for Classification in Imbalanced Datasets in the Presence of Noise	Chongomweru Halimu, **Asem Kasem***	*UTB	CIIS	17-19/10/2021	Virtual Conference
12	MM-AQI: A novel framework to understand the associations between urban traffic, visual pollution ,and Air pollution	Kazuki Tejima, **Minh-Son Dao***, Koji Zettsu	*NICT	IEA/AIE(rank B)	19-22/07/2022	Kitakyushuu、Japan
13	Federated Learning for Air Quality Index Prediction: An Overview	Duy-Dong Le, **Mohamed Saleem Haja Nazmudeen*,** Anh-Khoa Tran, **Minh-Son Dao**,** Viet-Tiep Mai and Nhat-Ha Su.	*UTB, **NICT,	KSE	19-21/10/2022	Vietnam

Scientific Contribution: International Conferences, Journal, Book Chapters, and Workshop Proceedings

No	Conference Paper Titles	Author names	Affiliation	Conference name	Date	Venue
14	Aedes Aegypti Egg Morphological Property and Attribute Determination Based on Computer Vision	Cherry R. Gumiran; Arnel C. Fajardo; Ruji P. Medina; Minh. S. Dao; Betchie E. Aguinaldo	ISU, NICT	IEEE ICSIP	20-22/07/2022	Virtual Conference
15	Towards Efficient Discovery of Periodic-Frequent Patterns in Dense Temporal Databases Using Complements	Pamalla Veena, Sreepada Tarun, R. Uday Kiran, **Minh-Son Dao***, Koji Zettsu, Yutaka Watanobe, Ji Zhang:	*NICT	DEXA (rank B)	22-24/08/2022	Virtual Conference
16	Adaptive Learning Models for Getting Insights into Multimodal Lifelog Data	Phuc-Thinh Nguyen**, Mohamed Saleem Haja Nazmudeen*** and **Minh-Son Dao****	*UTB, **NICT	KSE	19-21/10/2022	Vietnam

No	Journal Paper Titles	Author names	Affiliation	Journal name	Publisher	Info
1	Research on Traffic Congestion Detection from Camera Images in a Location of Dalat	Nguyen Thi Luong	DLU	Dalat University Journal of Science	Dalat University, Vietnam	11(4), pp. 63-75 (2021)
2	Convolution Recurrent Neural Network for Daily Forecast of PM10 Concentrations in Brunei Darussalam	Aziz E.N., Kasem A., Haji Suhaili W.S., Zhao P.	UTB, NICT	Chemical Engineering Transactions	The Italian Association of Chemical Engineering	83, pp.355-360 (2021)
3	A novel ensemble method for classification in imbalanced datasets using split balancing technique based on instance hardness (sBal_IH)	Halimu Chongomweru, **Asem Kasem***	*UTB	Neural Computing and Applications	Springer	33, p.11233-11254 (2021)
4	Assessing bikeability with street view imagery and computer vision	Koichi Ito, **Filip Biljecki***	*NUS	Transportation Research Part C: Emerging Technologies	Elsevier	132, pp. (2021)
5	[*Book Chapter*] Insights for Urban Road Safety: A new Fusion-3DCNN-PFP Model to Anticipate Future Congestion from Urban Sensing Data	**Minh-Son Dao***, R.Uday Kiran, Koji Zettsu	*NICT	[*Edited book*] Periodic Pattern Mining: Theory, Algorithms, and Applications	Springer	pp. 237-263 (2021)

No	Workshop proceedings	Organizer Names	Affiliation	Conference name	The date of the conference	The venue of the conference
1	ICDAR'21: Intelligent Cross-Data Analysis and Retrieval	**Minh-Son Dao* ,** Michael Alexander Riegler , Duc-Tien Dang-Nguyen , Cathal Gurrin , Minh-Triet Tran , Thanh-Binh Nguyen	*NICT	ACM ICMR	16-19/09/2021	Virtual Conference
2	ICDAR'20: Intelligent Cross-Data Analysis and Retrieval	**Minh-Son Dao***, Morten Fjeld, Uraz Yavanoglu, **Filip Biljecki****, Mianxiang Dong	*NICT,**NUS	ACM ICMR	26-29/10/2020	Virtual Conference
3	MediaEval2021: Insight for Wellbeing	**Asem Kasem*, Minh-Son Dao**, Effa Nabilla Aziz***, D.T. Dang-Nguyen, Cathal Gurrin, M.T. Tran, T.B. Nguyen, **Wida Suhaili***	*UTB, **NICT	MediaEval	13-15/12/2021	Virtual Conference

Scientific Contribution: Social Impacts

1. Open public datasets:
   - Haze Transboundary: ASEAN Archives of weather and air pollution data.
   - Environment-visual datasets in Dalat city, Vietnam: Dalat city archive of weather, air pollution, and CCTV videos.
   - Aedes Aegypti Egg Images captured in Cauyasan City
2. Challenges:
   - MediaEval 2021 Insight for Wellbeing: Cross-Data Analytics for (transboundary) Haze Prediction (https://multimediaeval.github.io/editions/2021/tasks/wellbeing/)
     - 10 teams registered
3. Conference special session and workshops:
   - ICMR2020 and 2021 workshop on Intelligent cross-data analytics and retrieval (ICDAR) (https://www.xdata.nict.jp/icdar_icmr2021/index.html)
     - 17 submitted papers, 5 accepted regular papers, and 4 short papers.
     - 1 keynote , 1 invited talk, and 1 panel.
   - MMM2022 Special session on Multimedia Analytics for Contextual Human Understanding (MACHU) (http://mmm2022.org/ssp.html#machu)
4. Towards Winning National Projects
   - ISU: C-DEWS Community Dengue Early Warning System
     - Budget: 100K (USD)
     - Time: 2 years
5. Spin-off Company
   - Original Organization: DLU, Vietnam
   - Domain: IoT network for Urban Environment-Mobility Assessment

Conclusions

1. **Scientific and technological**
   - Researched and developed several topics and methods that contribute to the success of the project including haze transboundary prediction on ASEAN area data, interactive incident retrieval in dashcam videos, periodic frequent patterns mining, air pollution estimation and prediction using images, and safe route navigation planning.
2. **Application (or system) development**
   - MM sensing system (including MM-trafficEvent, MM-trafficNet, MM-mobility-AQI)
   - A risk-avoidance navigation system
   - Dengue Early Warning System architecture
   - APIs library and manual instruction website
3. **Experiments including field testing**
   - Evaluated (transferred) haze transboundary prediction models in ASEAN dataset
   - Evaluated MM-trafficEvent in Tokyo, Dalat city, and youtube datasets
   - Evaluated MM-mobility-AQI in Tokyo, Dalat city, and India datasets.
   - Evaluated risk map in Dalat city dataset
   - Evaluated MM-trafficNet in Dalat city dataset
   - Evaluated Fed xData, bikeability with street view
4. **Datasets**
   - Tokyo environment-lifelog dataset
   - Dalat city environment-lifelog-tourism-CCTV dataset
   - Cauayan city Dengue dataset
   - ASEAN environment-air pollution dataset

Future works

1. Scientific and technological
   - Continue ongoing scientific and technological purposes mentioned above by singing CRA and MOU between project members
   - Research and Develop more topics and methods focusing on improving the smart of data, the flexible ability of Reusable, Sharable, and Transferable of the system.
2. Application (or system) development
   - Smart environmental Tourism and Sustainable Mobility for Dalat city, Vietnam
   - Smart Dengue Early Warning System for Cauayan City, Philippines
   - Smart Outdoor Activities for Singapore
   - Transboundary Air-Pollution Forecasting for ASEAN countries
   - Open dataset for open science data community
   - Completed xDataPF – xDataEdge and MM sensing system with high privacy protection
3. Experiments including field testing
   a) Field experiment of environmental quality data collection by residents using MM sensing and customization of short-term prediction of environmental quality of tourist spots (Dalat City, Vietnam) and environmental health (Cauayan City, Philippines) using collected data.
   b) Benchmarking to attract the attention from industry-academy-government on environment-human topics, expected to be organized in NUS