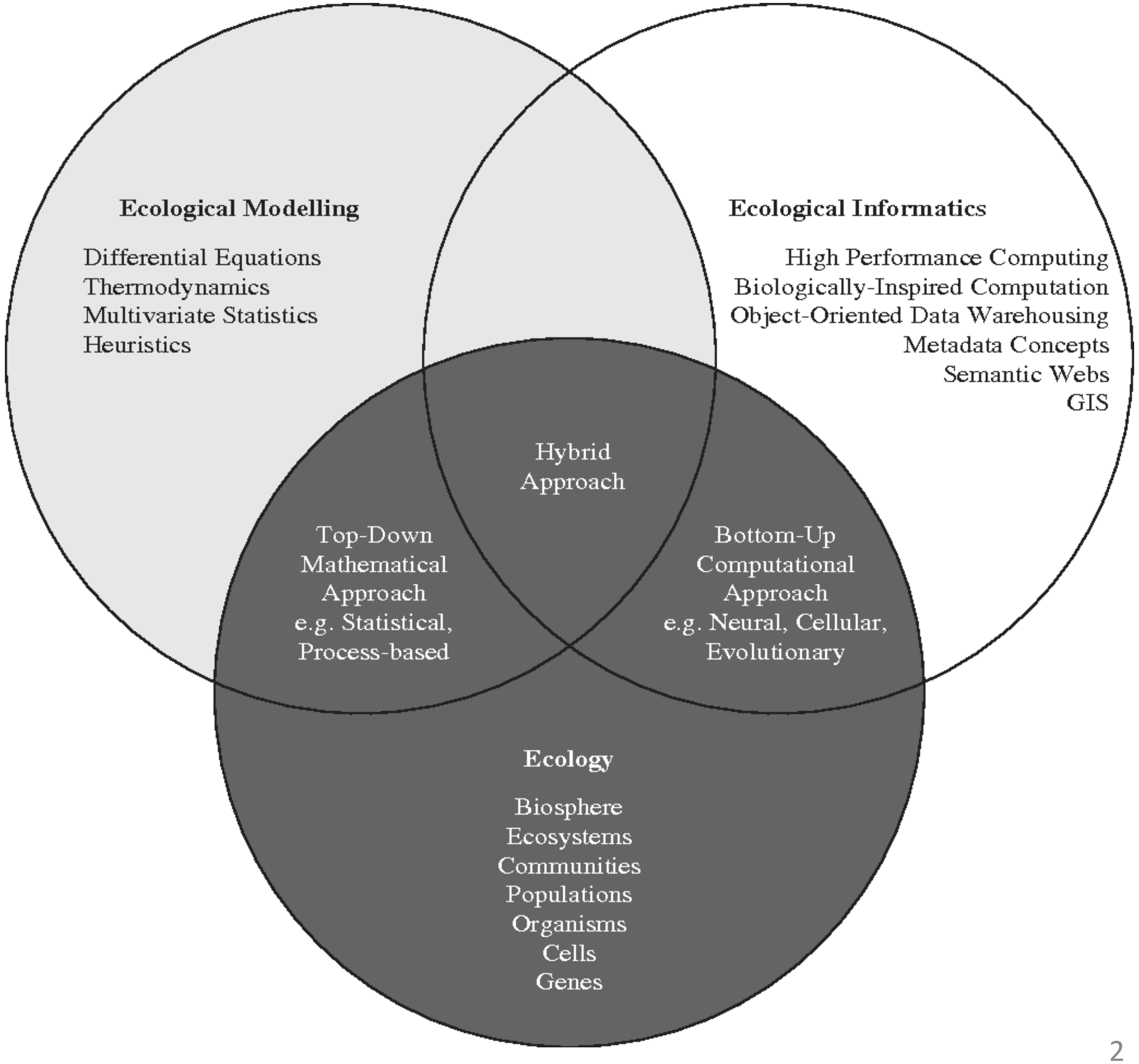
PRAGMA 39 Workshop

Deep Learning Applications in Ecological Informatics

Assoc. Prof. Dr. Nurul Hashimah Ahamed Hassain Malim
Universiti Sains Malaysia
22 June 2023

Emerging discipline that considers the data intensive nature of

- ✓ the valuable information content of ecological data
- ✓ the need to communicate results and decision making
- ✓ for research, conservation and resource management (Recknagel, 2017).



Challenges of Harmful Algal Bloom (HAB) Prediction

- Water resources have been reported to be polluted by **Harmful algal bloom (HAB)** that can cause harm.
- Since algae communities comprise of
  - Various species
  - Differ in nonlinear ways
  - Complex
  - Dynamic growth
- **Hard to analyse** and are **not well understood**, resulting in **unreliable predictive models**
- The dynamic growth of algae, which can vary on short timescales (e.g., hours to days) has made identifying the condition that favours HABs a major research effort.

Challenges of Harmful Algal Bloom (HAB) Prediction (Cont.)

- Existing ecological studies, especially those on the algae population are lacking in several aspects.
- To achieve robust predictive modelling of algal growth, several issues must be highlighted and addressed, for example,
  - (i) **the features must be mapped to the dynamic issues of algae ecology** and
  - (ii) **a suitable algal growth predictive modelling must be found, particularly to tackle dynamic algae for coastal studies**. Because prediction has been mostly done in rivers and lakes
- Addressing the problems through the features (water parameter) level and algorithm level might help to achieve the main aim of this research **of solving the dynamic issues**

Challenges of Harmful Algal Bloom (HAB) Prediction (Cont.)

- With the current advancement of the Internet of Things (IoT), the process of monitoring and profiling water quality and eutrophication mitigation can be facilitated using sensors
- We have came up with the end-to-end solution including:
  - assembly and integration of sensors,
  - data acquisition
  - and predictive modelling

MONITORING AND PROFILING WATER QUALITY FOR GROWTH OF ALGAE

Kit Development Phase

Sensor Integration

- pH
- EC
- DO
- Turbidity
- Temperature
- TDS

ARDUINO

Maintenance

- Calibration each probe
- Data Logging to SD
- Power management- setup battery & Solar
- Housing Preparation

SCHOOL OF COMPUTER SCIENCES
UNIVERSITI SAINS MALAYSIA

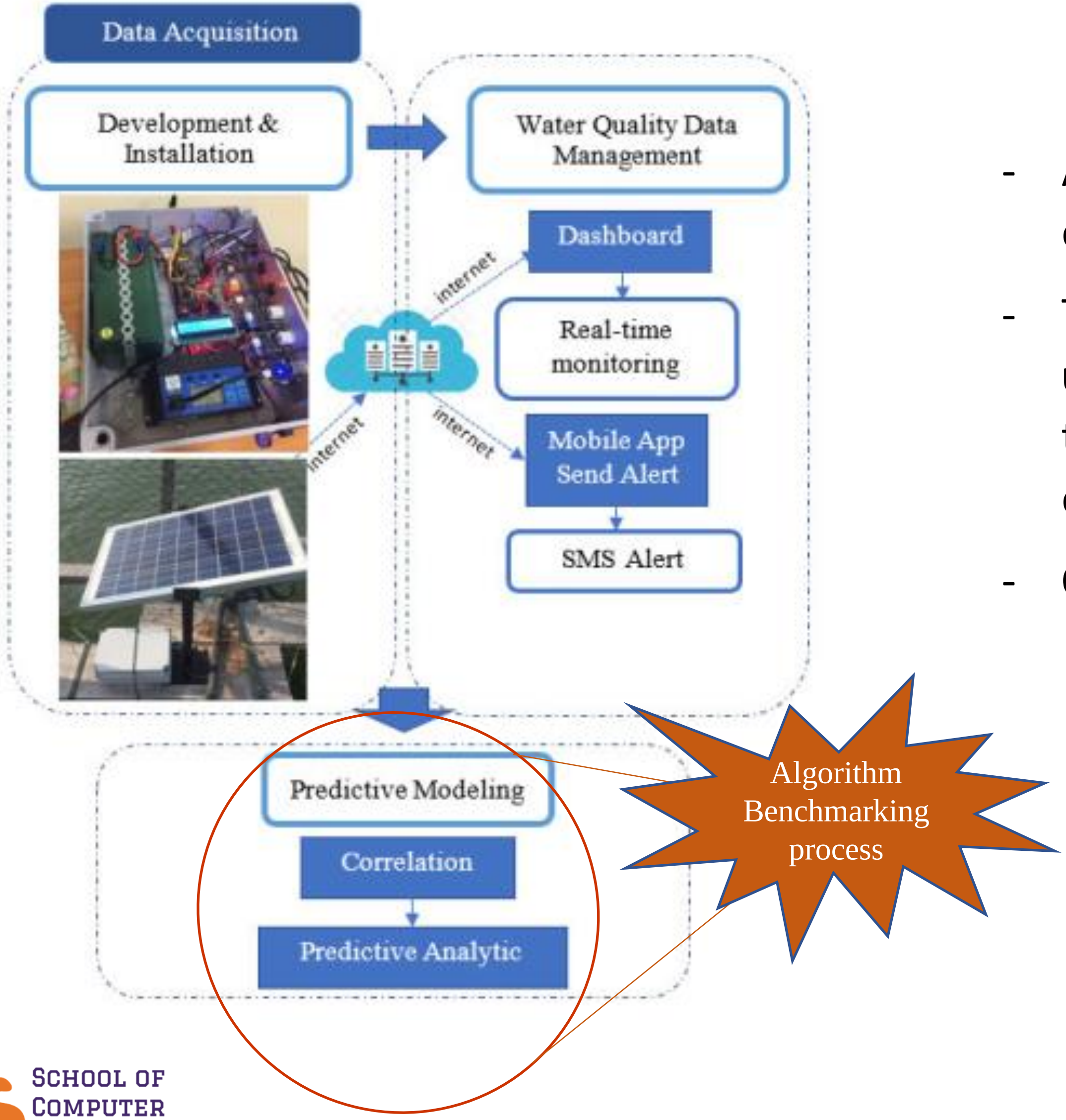
MONITORING AND PROFILING WATER QUALITY FOR GROWTH OF ALGAE

- Automated WQMS for aquaculture enable
  - **The industry to decrease catastrophic losses**
  - **Lower production cost and**
  - **Enhance product quality**

- Inexpensive sensors ensure
  - **Successful integration of obtaining the needed data**
  - **Predictive modelling development**

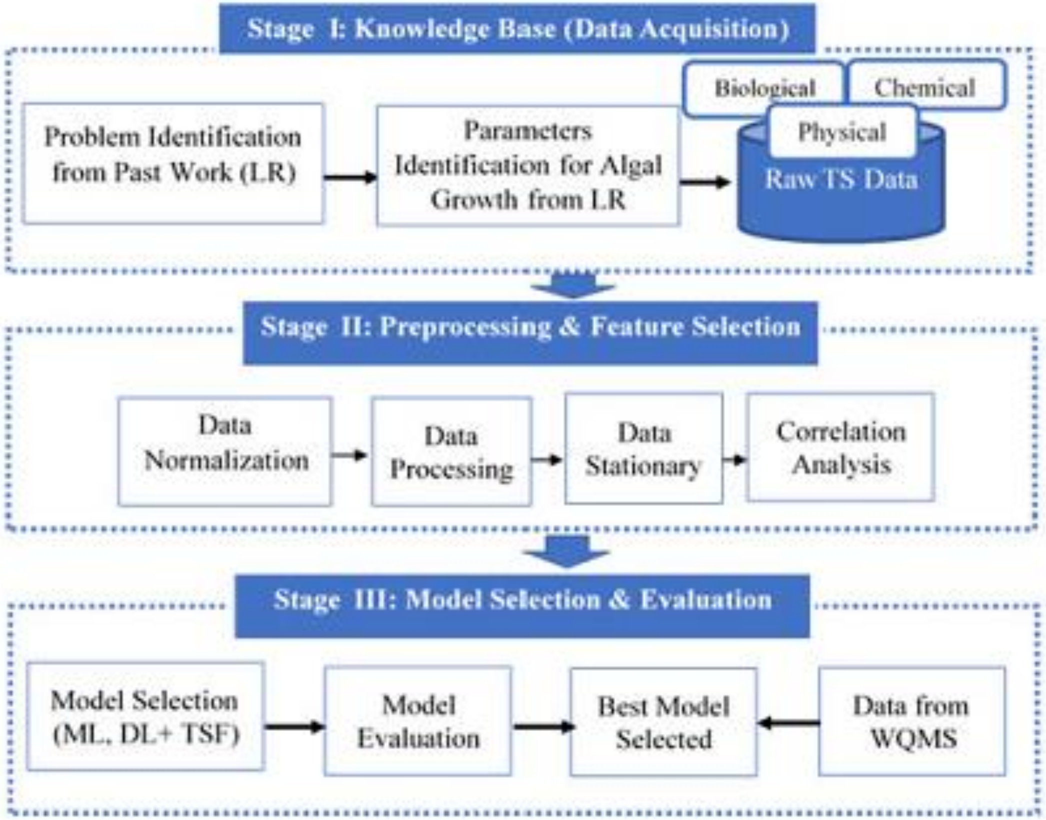
- Inspired from a Smart River Monitoring System for river concept paper seawater monitoring system was studied and proposed in our previous work

ALGAL BLOOM PREDICTIVE MODELING



- As a revision to our previous work, current work presents an enhanced and more detailed predictive modelling
- This work **presents proof that selecting the right features and utilising time series with deep learning** are much better for tackling the issues of highly non-linear and dynamic algae ecological data.
- **Overall steps** in conducting this research
  I. Identifying important factors for algal growth
  II. Review monitoring and profiling past work
  III. Review data-driven predictive model past work

PROPOSED PREDICTIVE MODELING FRAMEWORK



Commonly used parameters in algal blooming monitoring or prediction

Abbreviation	Variable	Factor Category
Chl-a	Chlorophyll-a	Biological Factor
BC	Bloom Cases (Incident)	**(BF)**
SGR	Specific Growth Rate	
WT	Water Temperature	
Salin	Salinity	
DO	Dissolved Oxygen	
Turb	Turbidity	
pH	pH	
SD	Secchi Disk Depth	Physical Factor
SS	Suspended Solid	**(PF)**
DC	Depth Code	
FI	Freshwater Inflow	
EV	Estuarine Velocity	
SRT	Salinity Recovery Time	
TIN	Total Inorganic Nitrogen	
PO_4	Orthophosphate	
TP	Total Phosphorus	
TN	Total Nitrogen	
AN	Ammonia Nitrogen	Chemical Factor
NO_2-N	Nitrite Nitrogen	**(CF)**
NO_3-N	Nitrate Nitrogen	
COD	Chemical Oxygen Demand	
Si	Silica	
Hg	Mercury	
Pb	Lead	
Zn	Zinc	
Al	Aluminium	
Rf	Rainfall	
T_{min}	Minimum Temperature	
T_{avg}	Average Temperature	Meteorological
T_{max}	Maximum Temperature	Factor
Hum	Humidity	**(MF)**
SR	Daily Solar Radiation	
WS	Daily Average Wind Speed	

Commonly used parameters to measure water quality

Water Quality Parameter	Abbreviation	Unit
Chlorophyll-a	Chl-a	mg/L
Secchi Disk Depth	SDD	m
Temperature	T	°C
Coloured Dissolved Organic Matters	CDOM	mg/L
Total Organic Carbon	TOC	mg/L
Dissolved Organic Carbon	DOC	mg/L
Total Suspended Matters	TSM	mg/L
Turbidity	TUR	NTU
Sea Surface Salinity	SSS	PSU
Total Phosphorus	TP	mg/L
Total Nitrogen	TN	mg/L
Orthophosphate	PO_4	mg/L
Chemical Oxygen Demand	COD	mg/L
Biochemical Oxygen Demand	BOD	mg/L
Electrical Conductivity	EC	Ms/cm
Ammonia Nitrogen	NH_3-N	mg/L

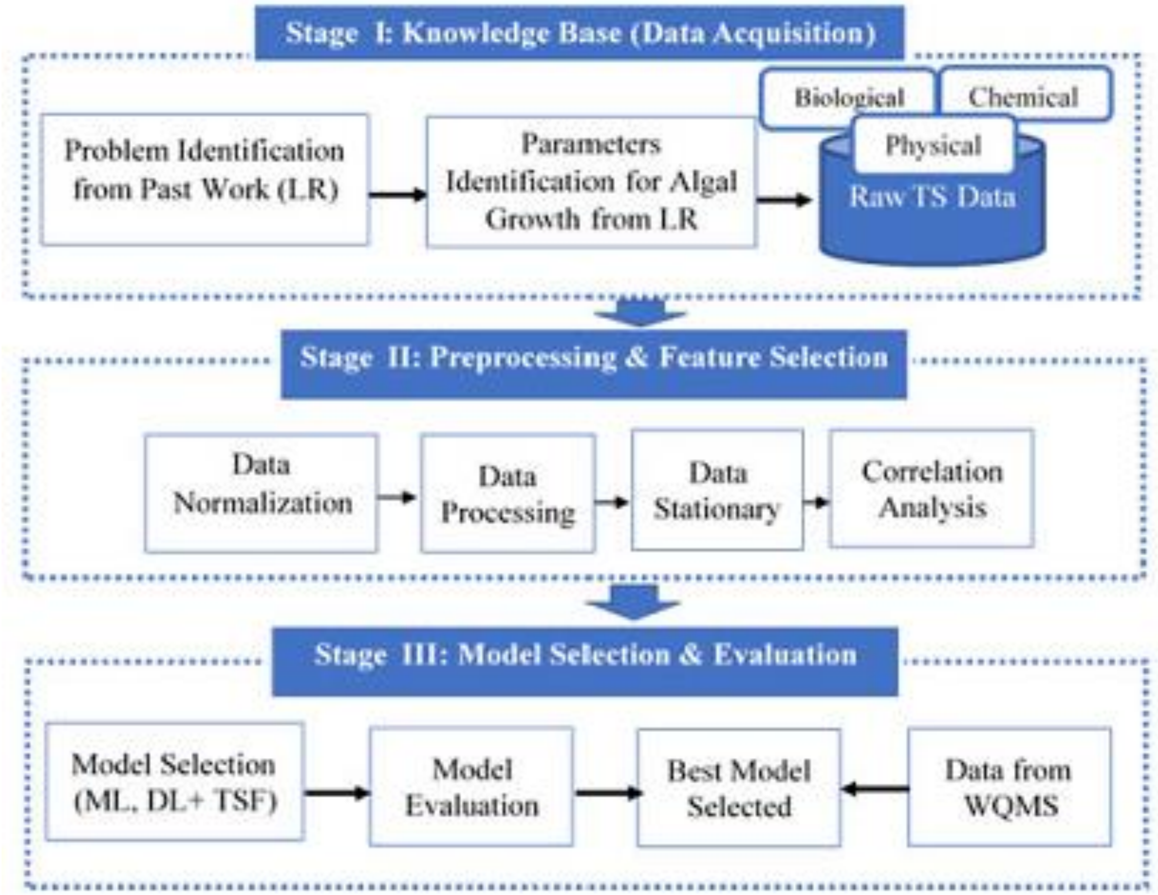
KNOWLEDGE BASE-DATASET DESIGN

- Based on past research on identifying important algal growth factors previously, investigation on PF (refer to Table 2) such as turbidity, DO, and other vital factors is still lacking.
- Hence, more features under the PF were included.
- Additional physical features common in water quality studies, such as salinity, turbidity, pH, suspended solids, and total nitrogen along with some CF variables were also included

Additional dataset design and description

Variable		Category
Chl-a	Chlorophyll-a	Biological (BF)
Salin	Salinity	
DO	Dissolved Oxygen	
Turb	Turbidity	Physical (PF)
pH	pH	
SD	Secchi Disk Depth	
SS	Suspended Solid	
Wtemp	Water Temperature	
TIN	Total Inorganic Nitrogen	
PO_4	Orthophosphate	
TP	Total Phosphorus	Chemical (CF)
TN	Total Nitrogen	
AN	Ammonia Nitrogen	
NO_2-N	Nitrite Nitrogen	
NO_3-N	Nitrate Nitrogen	
Si	Silica	

IDENTIFYING IMPORTANT FACTORS OF ALGAL GROWTH



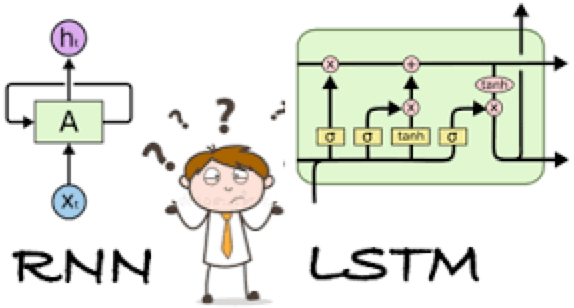
- Pre-processing that include:
- **Min Max Normalisation** is a rescaling of data from the original range so that all the values are within the range of 0 and 1.
- **Linear interpolation** method - impute missing data
- time-series data was framed as a supervised learning problem
- Make data stationary using **Augmented Dickey Fuller (ADF)** for smooth forecasting
- Feature Selection using **Correlation Analysis technique** - investigating the relationship and measuring the strength between two quantitative, continuous variables to represent their interdependencies.

DATA-DRIVEN ALGAL GROWTH PREDICTION METHODS

- **Using data-driven approaches:**
  - **Machine learning**
  - **Deep learning and**
  - **Time series forecasting**

- Limitation basic machine learning models
  - Unable to extract features of multi-factor timing data
  - Not reflect the temporal characteristics of the data
- Alternative approach:
  - Deep learning and time series

MODEL SELECTION-DEEP LEARNING (LONG SHORT-TERM MEMORY)



- RNN and LSTM **has shown outstanding performance to capture non-linear and temporal behaviors**, but it has not been applied on coastal datasets.
- LSTM neural network, a variant of RNN, was initially introduced to solve the vanishing gradient issue, which causes training divergence in RNN
- Like RNN, **LSTM is very capable of capturing the dynamic features via cycles in the graph**.

COMPARISON OF PREDICTION MODEL

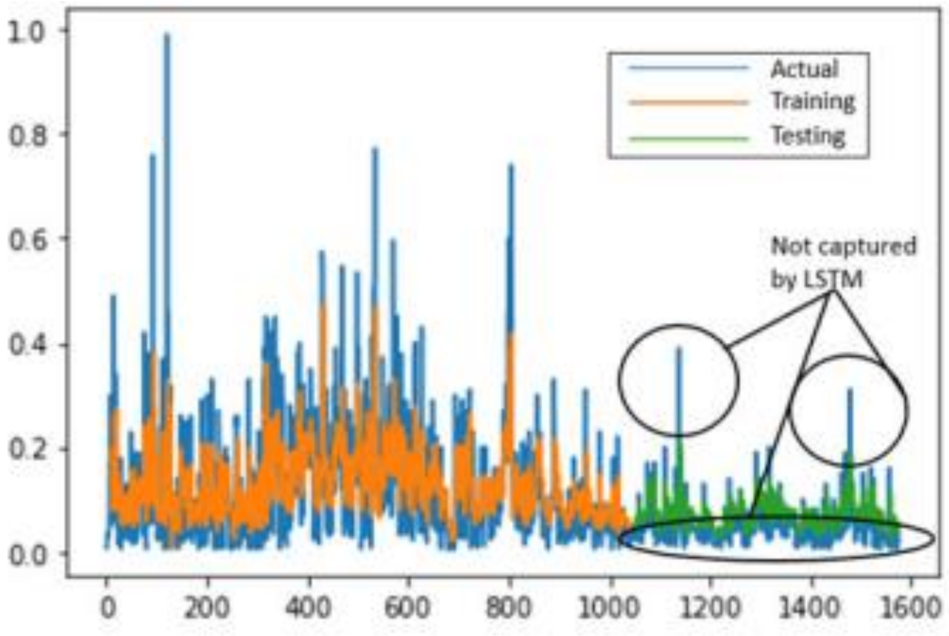
Comparison model performance evaluation for testing data

Method	MAE	RMSE	MSE
SVM	0.4772	0.5923	0.3508
DT	0.4840	0.5940	0.3528
RF	0.4453	0.5686	0.3233
MLR	0.4477	0.5632	0.3171
ANN	0.5607	0.6359	0.4044
TSP	0.4772	0.5923	0.3508
RNN	**0.0594**	**0.0696**	**0.0048**
DNN	**0.0319**	**0.0440**	**0.0019**
LSTM	**0.0256**	**0.0360**	**0.0013**

Comparison model performance of our approach and LR

Author(s)	Method	Source	MAE	RMSE	MSE
[23]	LSTM	River	NP	0.0486	NP
[28]	LSTM	River	NP	7.67	NP
[25]	Merge-LSTM	River	NP	0.0459	NP
[67]	DA-RNN	Coastal	0.790	1.269	NP
[12]	SVM	Coastal	0.926	1.583	NP
Ours	**SVM**	**Coastal**	**0.477**	**0.592**	**0.351**
Ours	**RNN**	**Coastal**	**0.091**	**0.083**	**0.008**
Ours	**DNN**	**Coastal**	**0.032**	**0.044**	**0.002**
Ours	**LSTM**	**Coastal**	**0.026**	**0.036**	**0.001**

*NP=Not provided



Model fitting of LSTM

DISCUSSION

FINDINGS

LSTM is a type of neural network with powerful nonlinear fitting ability.

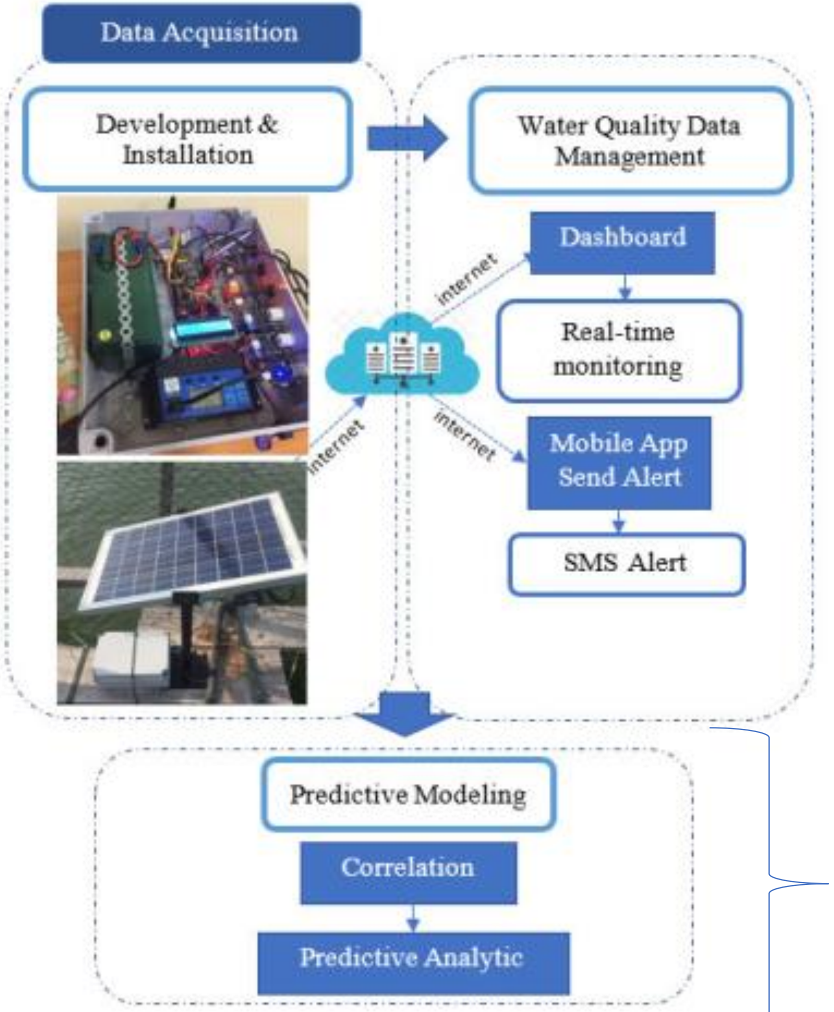
PF, BF, and CF factors are enough as inputs to the development of predictive modelling.

FUTURE

More attention on predicting the sudden high peak that is not captured by LSTM

Improvement through the selection of significant features is the best in handling non-linear, uncertain and dynamic data

EVALUATION OF A REAL CASE STUDY USING THE PREDICTIVE MODEL




(a) (b) (c)

Successfully predicted the next ten days potential readings for turbidity sensor

CONCLUSION

Algae issues that are highly **nonlinear** and **uncertain**, robust predictive modelling that tackles from the end-to-end process is necessary.

Selecting the right features are crucial in tackling the dynamic issues, and from the results, the algae ecology is dependent on the number and types of the features.

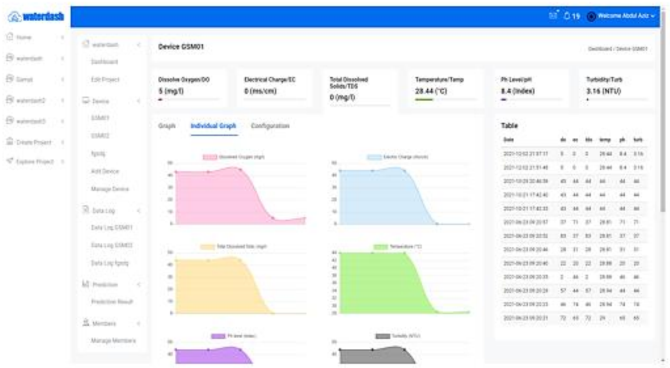
LSTM with the right features outperformed the other methods and grasped the temporal behaviour and tackled the dynamic issues.

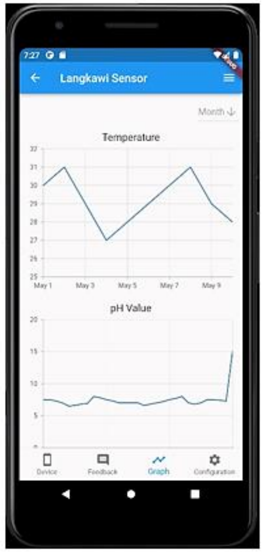
Besides, even though during this study **excluded meteorological factor, and more chemical and physical factor were included,** this study outperformed the other studies.

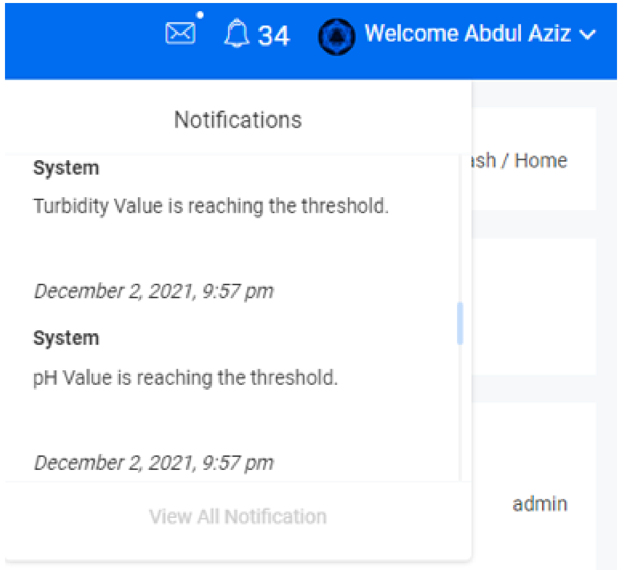
UNIVERSITI SAINS MALAYSIA

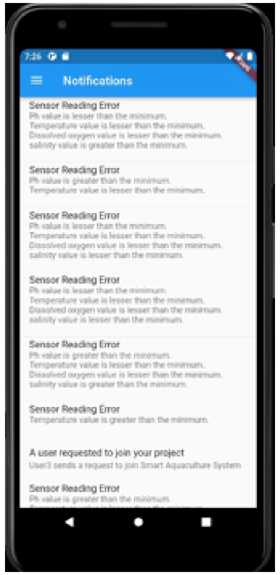
APEX™

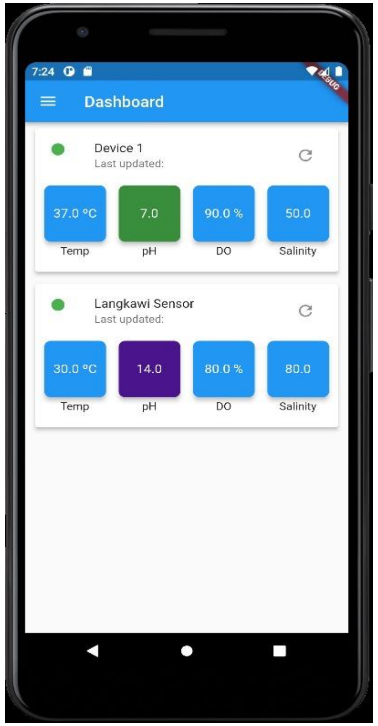
WATERDASH

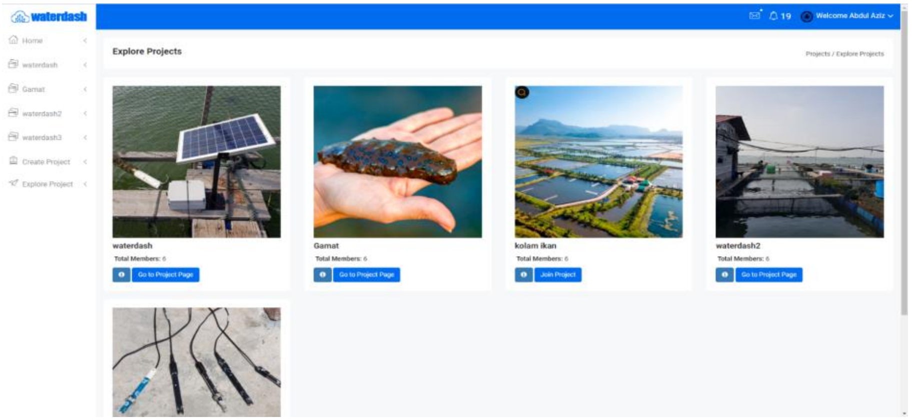


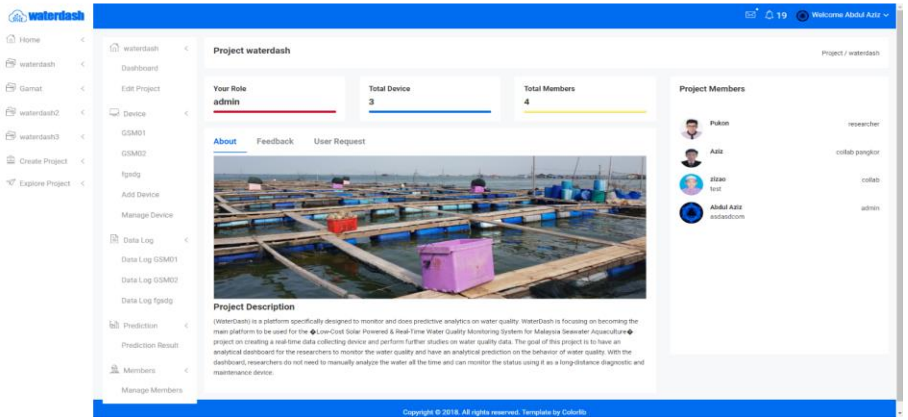










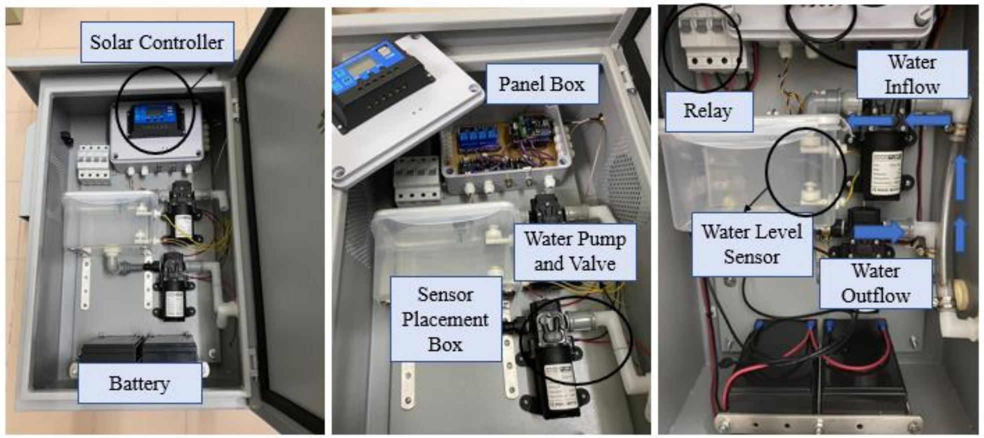


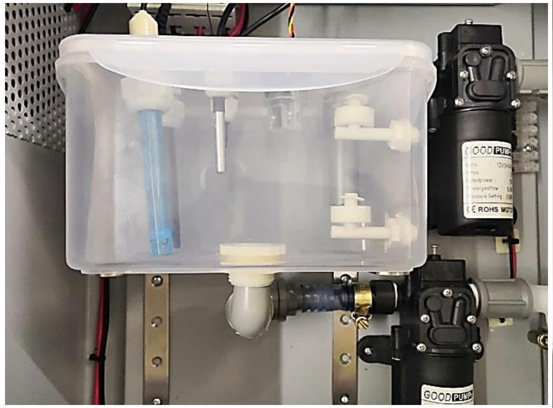
waterdash.cs.usm.my

SCHOOL OF COMPUTER SCIENCES
UNIVERSITI SAINS MALAYSIA

https://www.youtube.com/watch?v=N_gUurBJQ-c

MORE PICTURES





Our setup at fish plant (sea water)

Our setup at CEMACS USM – sea cucumber monitoring for indoor breeding

UNIVERSITI SAINS MALAYSIA

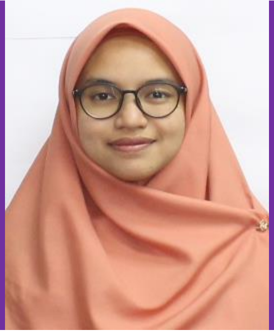
ACKNOWLEDGEMENT

GRANTS

Project Title: Sensor-based profiling and predictive analytics of solar flux and water quality on the seawater aquaculture
Fund: Transdisciplinary Research Grant (TRGS/1/2018/USM/01/5/4-203.PKOMP.67612) by Ministry of Higher Education Malaysia
Project Title: Smart and Precision Gamat Aquaculture
Fund: Transdisciplinary Research Grant (TRGS/1/2020/USM/02/2/1) by Ministry of Higher Education Malaysia

PUBLICATIONS

NAP Rostam, NHAH Malim, R Abdullah, AL Ahmad, BS Ooi, DJC Chan, A Complete Proposed Framework for Coastal Water Quality Monitoring System With Algae Predictive Model (2021) IEEE Access 9, 108249-108265. Nur Aqilah Paskhal Rostam, Nurul Hashimah Binti Ahamed Hassain Malim, Rosni Binti Abdullah @Mustafa, Development of a Low-Cost Solar-Powered & Real-Time Water Quality Monitoring System for Malaysia Seawater Aquaculture: Application & Challenges (2020) Proceedings of the 2020 4th International Conference on Cloud and Big Data Computing in ACM International Conference Proceeding Series, :, 86-91.



Nur Aqilah Paskhal Rostam

PHD

Abdul Aziz Bin Abdul Hadi

FYP Graduate
Master

Muhammad Aizuddin

Master

SCHOOL OF COMPUTER SCIENCES
UNIVERSITI SAINS MALAYSIA

Thank you

nurulhashimah@usm.my

UNIVERSITI SAINS MALAYSIA

SCHOOL OF COMPUTER SCIENCES
UNIVERSITI SAINS MALAYSIA