Supplementary Materials: An Agile AI and IoT-Augmented Smart Farming: A Cost-Effective Cognitive Weather Station

Supplementary Part I

Table S1. Some of the most common challenges in IoT.

Dilemma	Description	Impact	Some Solutions
Energy	A situation where the system's performance and functionality are impacted due to energetic challenges (high power consu- mption or low energy efficiency)	short lifetime. Low performance. High insta- bility	WSN Clustering. Software and Hardware optimization. Low-power wireless protocols.
Security	An issue or a risk that is based on threats or vulnerabilities.	Data breach. System and services un-availability.	Data encryption. Decentralized systems. Usage of White/Black list authentication. Usage of VPNs.
Cost	High cost related to hardware, software, and li- censes	Limited use of solutions.	Usage of open-source technology. Usage of low-cost hardware. Migrate hardware based solution to software based services.
Scalability	The ability to have a system that are customized for different use cases and able to be upgraded for future requirements.	High costs. Slow time to market, where solutions need to be developed from scratch.	Usage of cloud-based solution. Adoption of remote accessibility and dynamic devices management approach. Flexible software architecture such as containerization.
Standards	Is the action to unify the diverse architectures to ensure interoperability.	Technical complexity. Integration challenges. Security and privacy. High cost for implementation.	Usage of standardized protocols such as HTTP, MQTT.
reliability	The ability that a system has to perform functions according to the specifications	Low performance	Implementation of OAM for monitoring and control. Components and services redundancy. Deployment of fault tolerance strategy.

Supplementary Part II

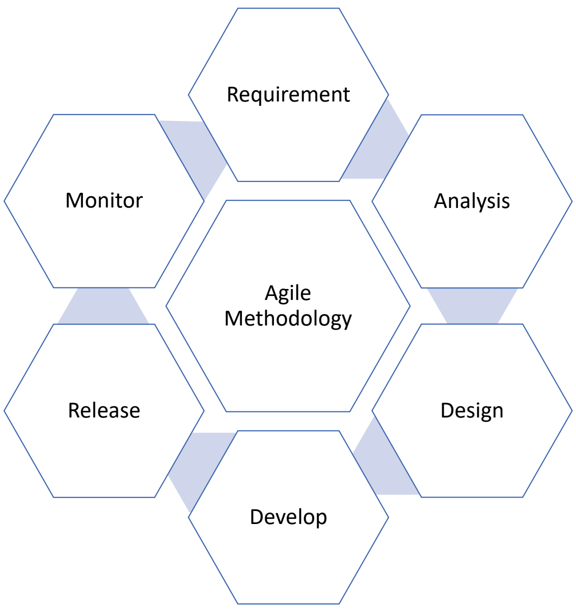

Figure S1. The key principles of Agile Methodology.

Supplementary Part III—Cost Effectiveness

Table S2. Capex investment.

Tool	Description	License	Environment	Type	Qty	Cost per Unit	Total
Raspbian	Operating System	Open-Source	Debian	OS	NA	Free of charge	NA
Docker	Container management	Open-Source	Debian	Software	NA	Free of charge	NA
Kubernetes	Container orchestration	Open-Source	Debian	Software	NA	Free of charge	NA
InfluxDB	database tool	Open-Source	Container	Software	NA	Free of charge	NA
Grafana	Visualization tool	Open-Source	Container	Software	NA	Free of charge	NA
OpenVPN	VPN management	Open-Source	Container	Software	NA	Free of charge	NA
Apache	web server	Open-Source	Debain	Software	NA	Free of charge	NA
Mosquitto	MQTT Server	Open-Source	Raspbian	Software	NA	Free of charge	NA
Raspberry PI	SoC	4B	Raspbian	Hardware	1	68$	68\$
Node32	SoC	32	Node	Hardware	1	10$	10\$
NRF24L01	wirless connection	-	2.4 GHz	Hardware	2	1$	2$
ESP8266	wirless connection	-	Wifi	Hardware	1	1$	1$
DHT11	Air Temp Humidity	-	-	Hardware	1	1$	1$
DHT22	Air Temp Humidity	Humidity	-	Hardware	1	3$	3$
FC-28	Soil moisture	-	-	Hardware	1	1$	1$
BMP085	Air Temp Humidity	Pressure	-	Hardware	1	1$	1$
PB200-286	Solar Radiation	-	-	Hardware	1	3$	3$
Arduino Micro	-	-	Arduino frim	Hardware	1	11$	11$
Arduino Uno	-	-	Arduino frim	Hardware	1	20$	20$
Anemometer	wind speed direction		-	Hardware	1	55$	55$
			Total Cost				176$

Table S3. Opex investment.

Item	Description	Unit	Qty	Value
Operation	Power consumption Raspberry	0.12$/KWh	5.47 kWh/1 month	0.7$
Operation	Power consumption Arduino Uno	0.12$/KWh	0.82 kWh/1 month	0.1$
Operation	Power consumption Arduino Mega	0.12$/KWh	0.67 kWh/1 month	0.1$
Domain	Website name registration	1.2$/month	1 month	1.2$
Web hosting	Cloud infrastructure	12$/month	1 month	12$
Maintenance	Equipment Maintenance	Open-Source	1 resource	Free of charge
Rent	Space for BS installation (Housing)	$/m^3$ /month	1 enclosure	Free of charge
		Total Cost		14.1$

Supplementary Part IV—LSTM Model

LSTM structure LSTM approach relies on the principle of gates to control the long-term sequence of information flux for a specific duration. The LSTM network Figure S2 can be seen as three gates, input gate i, output gate o, and a forget gate f :

- the input gate with its activation Vector i_t determines the information flux that should enter the cell,

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i). \quad \text{(S1)}$$

- the output gate with its activation Vector o_t determines the information flux that should pass to the next cell,

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o). \quad \text{(S2)}$$

- the forget gate with its activation Vector f_t determines the information flux that should be saved in the network's memory or should be forgotten,

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right), \tag{S3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \tag{S4}$$

$$h_t = o_t \odot \tanh(c_t). \tag{S5}$$

In our model, a Rectified Linear unit (ReLU) function Equation (S6) is used as an activation function of the deployed neurons:

$$\sigma(x) = \begin{cases} 0 \text{ for } x < 0, \\ x \text{ for } x \geq 0. \end{cases} \tag{S6}$$

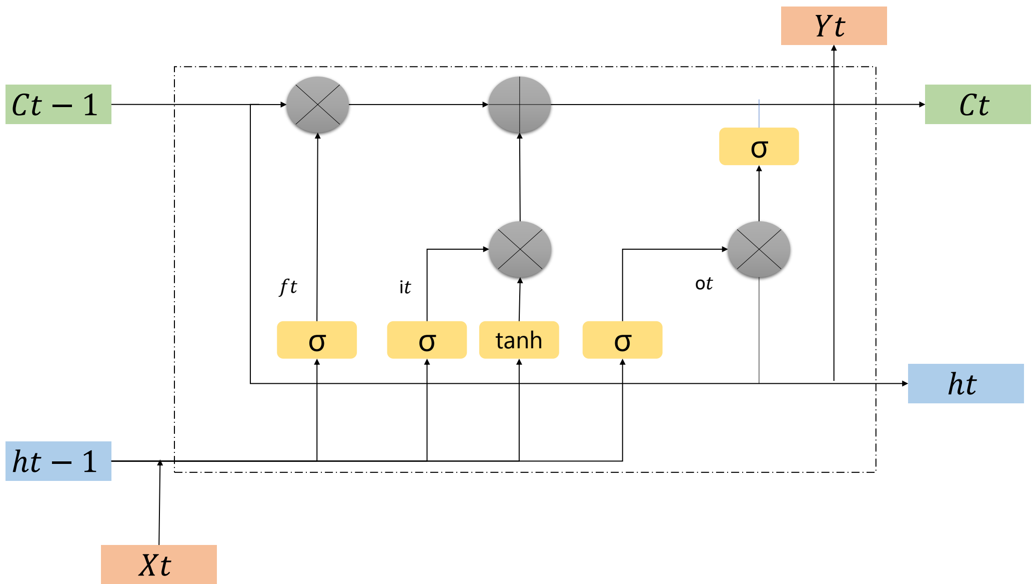

Figure S2. The structure of a single cell within LSTM.

The LSTM network learns from long and short-term features in the studied training dataset. Therefore, the usage of non relevant data sources, or an adequate learning and optimization process may lead towards a wrong or unclear predictions. Thus, in our work we implement the workflow depicted in Figure S3 to achieve the best optimum results.

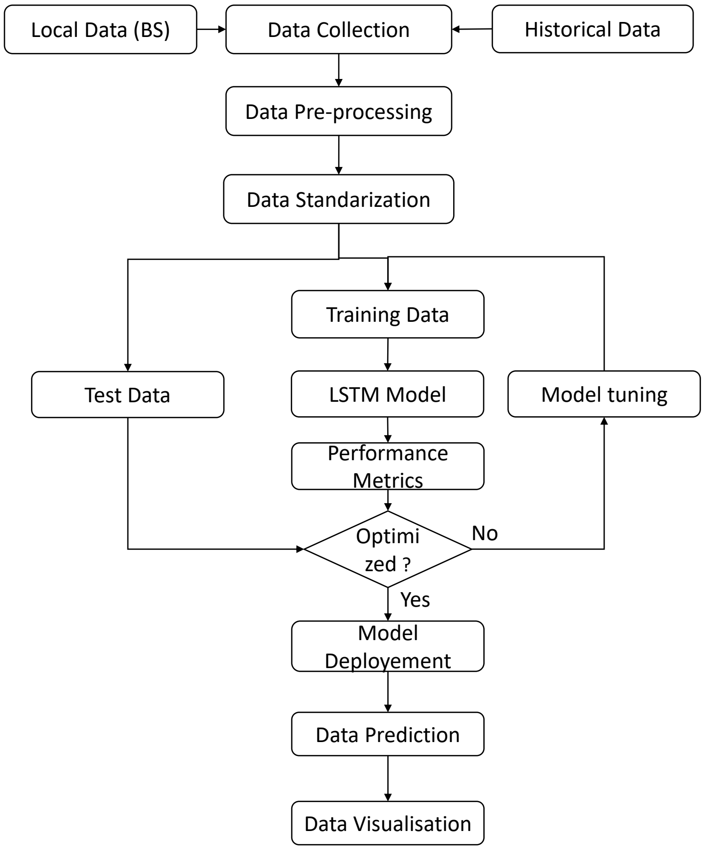

Figure S3. Model work flow of a single cell.

Supplementary Part V—Feature Correlation

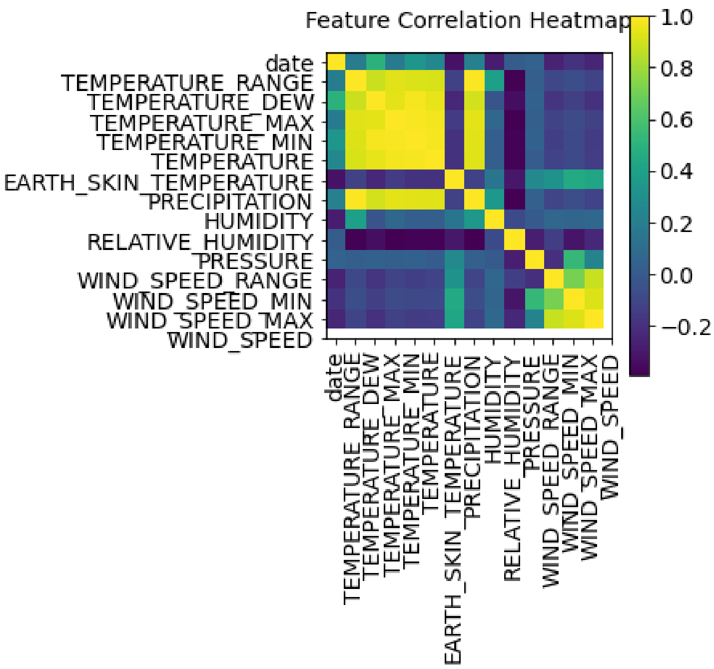

Figure S4. Feature selection using heat map method.

Supplementary Part VI—Model Validation

We plot the model results after each epoch which shows improvement of the relationship between prediction and real values throughout the iterations. The data correlation trending throughout the rounds is presented in Figure S5.

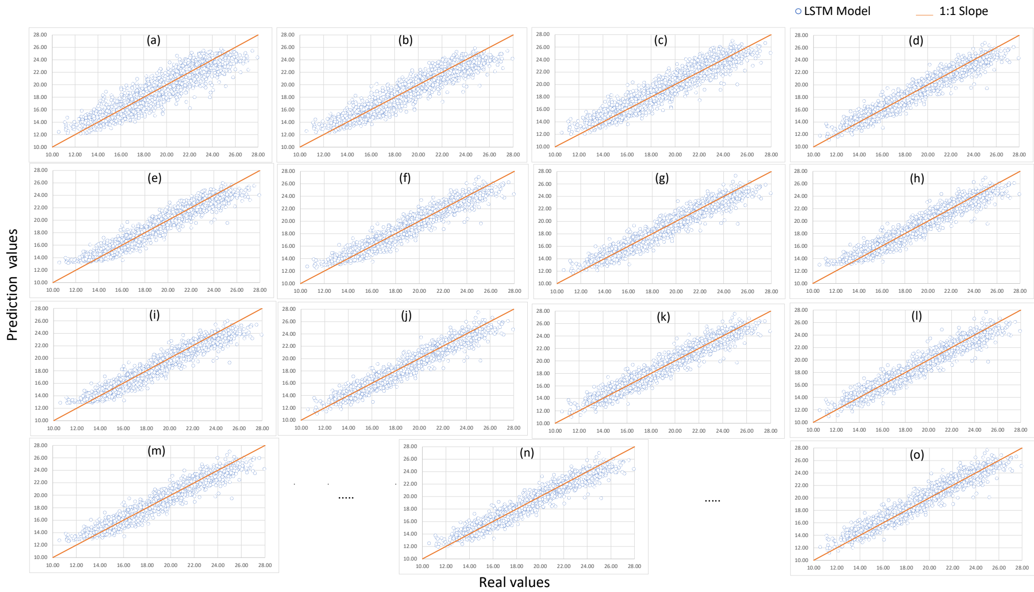

Figure S5. Scatter plot—Real values vs prediction values.

The daily residual values are illustrated in Figure S6 for 1 month, 1 year, and 8 years. Meanwhile, it is visible that there are some residual peaks that varies in worst case between 3.6 and −5.08.

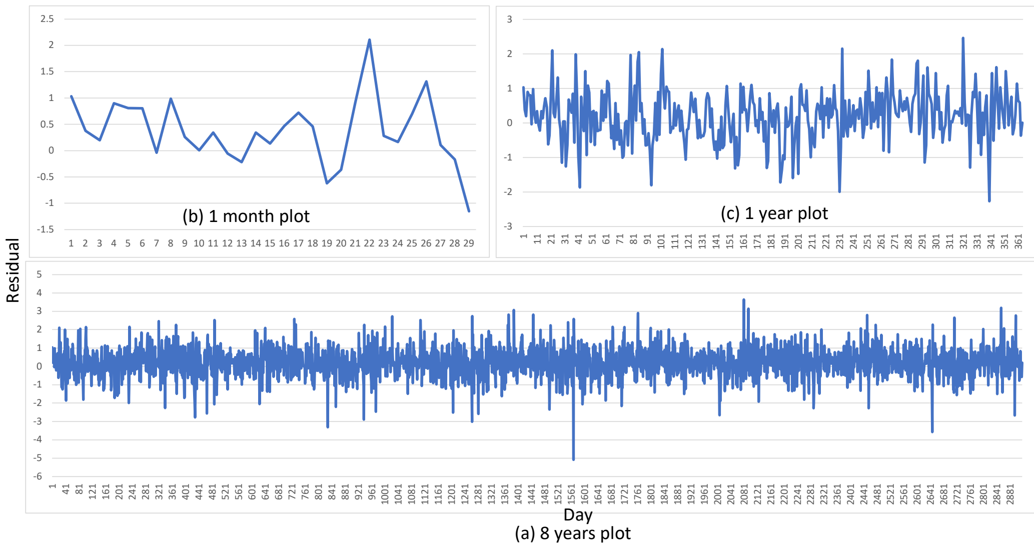

Figure S6. Residual plot—Muli-timeframe presentation. (**a**) daily Residual plot for 8 years (entire dataset). (**b**) daily Residual plot for 1 month. (**c**) daily Residual plot for 1 year.

Even though the outliers are clearly displayed on the scatter we can see in Table S4 that 51% of the prediction are within the range of [0 °C, 1 °C], 30% are within the range of [−1 °C, 0 °C], 12% of the records are within [1 °C, 2 °C], while only 7% of remaining records are beyond these ranges.

Table S4. Range and percentage of deviation records.

Deviation Range (°C)	Number of Records	Percentage
≤ -5	1	0%
-4–-3	3	0%
-3–-2	19	1%
-2–-1	113	4%
-1–0	881	30%
0–1	1497	51%
1–2	361	12%
2–3	37	1%
>3	5	0%
Total Records	2917	100%

Supplementary Part VII

In the Table S5 we present the major advantages and challenges that we detected during the analysis of different systems. We also categorize the existing platforms based on various metrics such as the validation environment. Additionally, we assess the systems' cost effectiveness based on deployed hardware and software. Finally, we categorized the usage of some features such as operation and maintenance, remote accessibility, and AI implementation.

Table S5. Comparison between the proposed platform and other existing platforms.

Ref	Advantages	Challenges	Validation	Cost	O&M	Security	Remote Access	AI
[20]	*Low-cost based agro-ecological management system. *Adoption WSN technology. *System based on lightweight protocols.	*High power consumption due to WSN. *Reliability issue when nodes go off	Simulation	Medium	No	No	No	No
[21]	*IoT-based platform for precision agriculture. *A real-time system for epidemic disease control. *Adoption of Solar panel for energy supply.	*Power supply will be impacted by the weather. *Unreliable central unit (based on Arduino). *Fully based on wired nodes.	Experimental	Medium	No	No	No	Yes
[22]	*A real-time monitoring system for farmers. *Implementation of an Expert Advisory System to improve crop productivity .	*High power consumption based on WSN adoption. *Reliability concern when nodes go off in WSN. Conceptual model need validation.	Simulation	Low	No	No	No	Yes
[23]	*An IoT-based low-cost system for smart irrigation. *The system is based on MQTT, and Neural Network (NN) for intelligent decision-making	*System complexity is high since it relies on wired nodes. *System only supports MQTT based nodes.	Experimental	High	No	No	No	Yes
[24]	*low-cost platform for smart farming *development of an user-friendly interface.	*System rely only on Power Bank for motherboard supply. *Conceptual model need to be validated on field	Experimental	High	No	No	No	No
[25]	*Proposition of an adaptive network mechanism for a reliable smart farming system. *The system implements LoRaWAN and IEEE protocols for transmission.	*Lack of end-to-end platform presentation and focus only on the transmission layer.	Experimental	Low	No	No	No	No
[26]	*An experimental analysis of energy harvesting for IoT devices. *Adoption of IoT devices with energy harvesting capabilities.	*Usage of basic limited HW for validation (Arduino Uno). limited resources (memory, computation, etc)	Experimental	Low	No	No	No	No

Table S5. *Cont.*

Ref	Advantages	Challenges	Validation	Cost	O&M	Security	Remote Access	AI
[27]	*An IoT system for precise ecological monitoring in agriculture domains. *Adoption of a friendly GUI for different users. *System open for upgrade.	*System high complexity because of lack of global GUI. *Gateway play role of central unit if node goes off system will be off.	Experimental	High	No	No	No	No
[28]	*Introduction of an architectural approach based on Farm Management Information System (FMIS)	*Conceptual model need implementation for validation.	Simulation	High	No	No	Yes	No
[29]	*An energy efficient weather station. *System focused on the algorithm optimization.	*System addresses only energy challenge.	High	No	No	No	No	No
[30]	*An UAV and IoT-based platform for smart farming. *Low-cost architecture to predict environmental data in large farms.	*High power consumption due to non-optimized communications. *lack or AI-based approach to enhance the overall performance.	Experimental	Low	No	No	No	Yes
[52]	*LoRa based architecture. *Low-cost system. *Modular framework. *Web-based visualization tool.	*System rely on single transmission protocol (LoRaWAN) *Power consumption	Experimental	High	No	No	No	No
[53]	*System support SMS-based notification. *Low-cost platform. *Usage of ML for crop prediction.	*A conceptual model need to be tested in production environment	Simulation	High	No	No	No	Yes
[54]	*Fuzzy and ANN based weather station for temperature forecasting. *Usage of solar panels.	*Power supply will be impacted by the weather.	Experimental	Medium	No	No	Yes	Yes
[55]	*Low-cost system for smart water irrigation. *Sel-reorginased WSN-based architecture. *Web-based application for users.	*High power consumption.	Experimental	High	No	No	No	No
[56]	*Low-cost system. *ML-based for rainfall prediction.	*Trivial communication model.	Simulation	High	No	No	No	Yes
[57]	*Platform based on deployment heterogeneous nodes. *Platform based on UAV.	*High power consumption due to the wireless communication.	Simulation	Low	No	No	No	No

Table S5. *Cont.*

Ref	Advantages	Challenges	Validation	Cost	O&M	Security	Remote Access	AI
[58]	*UAV-based framework for agriculture. *adoption of smart energy-harvesting model.	*System had some limitations like limited and short coverage. *Unreliable data transmission.	Experimental	Low	No	No	No	No
[59]	*An IoT-based AgroTech is proposed for urban farming *System System allows close monitoring of temperature, humidity, and soil moisture.*System's reliability us verified based on error modeling..	*Very challenging in power consumption.	Experimental	Low	No	No	No	No
AWS	*An agile AI and IoT-augmented weather station. *A real-time weather station for smart farming support multi-user multi-profile architecture. *A fully containerized-based architecture for continuous upgrade. *A system based on Plug-and-Play wireless nodes. *A multi-agent-based weather station. *System supports alarms, notification and automation. *Implementation of a sophisticated Operation and Maintenance dashboard for the BS control. *System offer high quality temperature forecasting based on LSTM. *System support Quality of Experience (QoE).	*Consideration of extension to support Lora, Sigfox, and NBIoT. *Need the design and implementation of solar panel-based solution for energy consumption. *real-time synchronization of the offline dashboard with the website: https://ensem-aws.tk/.	Experimental	Low	Yes	Yes	Yes	Yes