CONESCAPANHONDURAS2025paper83.pdf

Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477772068

Submission Date

Jul 31, 2025, 11:04 PM CST

Download Date

Aug 12, 2025, 2:48 PM CST

File Name

CONESCAPANHONDURAS2025paper83.pdf

File Size

1.3 MB

6 Pages

3,287 Words

19,572 Characters

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **16** Not Cited or Quoted 11%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **1** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 11% Internet sources
- 11% Publications
- 0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **16** Not Cited or Quoted 11%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **1** Missing Citation 1%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

11% Internet sources
11% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org	2%
2	Internet	journal.ilmudata.co.id	<1%
3	Internet	digitalcommons.unl.edu	<1%
4	Publication	Alfred Toku, Samuel Twumasi Amoah, Nelson Nyabanyi N-yanbini. "Exploring the ...	<1%
5	Publication	P Vineeth, T Ananthan. "Automated Hydroponic System using IoT for Indoor Far...	<1%
6	Publication	Amith Khandakar, I.M. Elzein, Md. Nahiduzzaman, Mohamed Arselene Ayari et al. ...	<1%
7	Internet	ijsrcseit.com	<1%
8	Internet	rmspagesq.stir.ac.uk	<1%
9	Publication	Md Rayhan Hossain, Md Nahiduzzaman, Md Rezwanul Karim, Md Tanvir Rahman,...	<1%
10	Internet	ia801700.us.archive.org	<1%

11	Publication	
Rúben El-Katib Penido, Rafael Christian Fonseca da Paixão, Laís Cristina Barbosa ...		<1%
12	Internet	
insightsociety.org		<1%
13	Internet	
sciencepublishinggroup.com		<1%
14	Publication	
Weiyue Xu, Tianyi Wang, Tianyu Ji, Qiong Su, Wei Chen, Changying Ji. "RepGhostC...		<1%
15	Internet	
www.biorxiv.org		<1%
16	Internet	
www.mdpi.com		<1%

Computational Tool for Statistical Analysis and Multivariate Forecasting in Aquaponic Systems Using Transformer-Based Deep Learning

1
1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—This work presents a hybrid computational tool for aquaponic systems that integrates automated statistical analysis, dynamic data visualization, and multivariate prediction through a dual output Transformer architecture. Using an IoT based monitoring system, critical variables such as pH, TDS, humidity, pond temperature, and ambient temperature were recorded. After robust preprocessing, a dual output Transformer model was trained to predict system behavior 24 hours in advance. The tool also provides interactive dashboards and statistical reports that facilitate interpretation and cultivation management, contributing to smart agriculture in aquaponic contexts.**

***Index Terms*—aquaponic systems, time series, deep learning, transformer, multivariate prediction, automated statistical analysis, Internet of Things, smart agriculture.**

I. Introduction

Various Latin American countries face critical challenges in the agricultural sector, arising from water scarcity, environmental degradation, and the urgent need to transition to sustainable production models. [1]. In this context, aquaponic systems [2], that integrate aquaculture with hydroponics in a closed environment [3], [4],They have emerged as an efficient and resilient solution, capable of producing food sustainably by recycling nutrients supplied by the fish and by optimizing water use. Despite their potential, the efficient operation of these systems remains limited in many local contexts, where manual methods for monitoring critical variables such as pH, water temperature, electrical conductivity, and light intensity still prevail. [5], [6]. These limitations reduce system efficiency and increase the risk of crop loss and higher fish mortality. In this scenario, the incorporation of automated monitoring technologies based on the Internet of Things (IoT) has enabled improved system observability and facilitated real-time environmental data collection. [7]–[9].

However, within the framework of Agriculture 4.0, simple monitoring is no longer sufficient. The evolution of these systems demands intelligent tools capable of processing large volumes of data, performing meaningful statistical analyses, and presenting results in a visual and comprehensible manner for producers. [10]. At a regional level, there is a lack of computational solutions specifically tailored to the requirements of aquaponic crops, integrating functions for the analysis, interpretation, and prediction of environmental data [11].

In response to this need, the present work proposes a comprehensive computational tool for advanced data analysis in aquaponic systems. This solution includes three fundamental components: an automated statistical analysis module that calculates key metrics such as minima, maxima, means, medians, and modes; an interactive visual dashboard that allows exploration of the temporal evolution of critical system variables; and a deep learning model based on a multivariate Transformer architecture, designed to predict the future behavior of environmental variables such as humidity, pH, TDS (total dissolved solids), pond temperature, and ambient temperature. The model incorporates a specialized branch to improve pH prediction, a variable particularly sensitive in these systems. The tool has been validated using real data obtained from an operational aquaponic environment with mixed crops of lettuce, tomato, eggplant, and cilantro, along with freshwater fish, demonstrating its utility for proactive decision-making, improved crop management, and the strengthening of sustainable agriculture in the region.

II. Methodology

A. Data Acquisition and Preprocessing

Data acquisition is a fundamental component in analysis and prediction systems, as it provides the basis for detecting relevant patterns, identifying anomalies, and understanding operational dynamics in complex environments. In the specific context of aquaponic crops, this task takes on particular importance due to the dynamic interaction among multiple environmental and physicochemical variables, which directly affect fish health and plant yield.

To capture these variables, an intelligent monitoring system based on IoT technologies was implemented, following the design proposed by Vásquez-Pineda et al [7],whose schematic is presented in the figure 1.This system employs distributed sensors to measure critical parameters such as pH, water temperature, TDS, relative humidity, and ambient temperature, continuously collecting data.

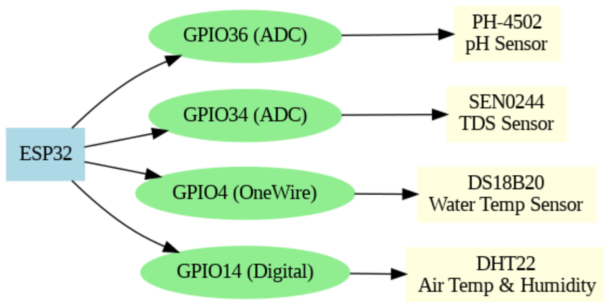

Fig. 1. IoT Monitoring System Connection Diagram

The data were recorded in a local database from November 6 to 11, 2024, resulting in the accumulation of approximately 10 million records. This considerable volume allowed for a detailed representation of the system's behavior under real operating conditions.

Prior to analysis, an extensive preprocessing procedure was applied. First, data cleaning was performed, which involved removing duplicate records and correcting outliers, defined based on valid physical ranges for each variable. Specifically, acceptable limits were established for pH and TDS values, and any measurements falling outside these ranges were replaced with nulls for subsequent handling. Next, linear interpolation techniques were used to fill missing values, followed by signal smoothing using median filters and Savitzky–Golay filters. This stage was crucial for reducing the noise inherent to field sensor measurements without losing the overall shape of the signals. Some variables, such as temperature and humidity, were also treated with centered moving averages to highlight local trends. Additionally, temporal variables such as time of day and day of year were transformed into cyclical representations using sine and cosine functions, enabling the capture of periodic and seasonal patterns present in the behavior of environmental variables. This encoding is especially useful in environments like aquaponic systems, where variable behaviors may follow natural daily or annual cycles. To ensure temporal coherence among different signals, all series were resampled to uniform 30 second intervals. This allowed precise synchronization between variables and prepared the data for multivariate analysis. Finally, the sequences were structured into fixed length input windows of 96 previous steps and prediction horizons of 2,880 future steps, thereby generating a dataset formatted appropriately to feed the Transformer model designed in this study.

The complete flow of this process is presented in Figure2, which summarizes the key stages from acquisition to the generation of datasets ready for statistical analysis and multivariate prediction using deep learning.

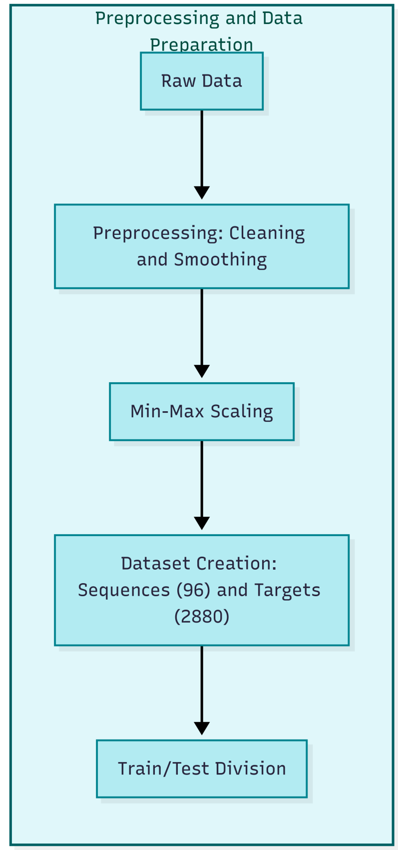

Fig. 2. Data Preprocessing Representation.

B. System Design and Architecture.

The proposed system was designed following a modular and scalable architecture aimed at facilitating the management, visualization, and analysis of sensor data in monitoring systems. The solution is structured into clearly differentiated functional layers: the user interface, data processing, and graphical visualization.

The main interface allows the user to create new projects, select input files in tabular format, and manage multiple variables related to physical or environmental processes. Once the data are loaded, the system automatically organizes the information and displays a main dashboard composed of individual cards summarizing key statistics for each variable such as maximum, minimum, mean, median, mode, and variance. In addition to individual analysis, comparative visualizations are enabled via side panels, where the user can observe relationships between pairs of variables. These charts allow identification of patterns, correlations, or anomalous behaviors over time through an intuitive, date navigable interface. The system's graphic design is optimized for clear representation of time series and provides an interactive visual analysis experience.

In the Figure 3 It illustrates the overall system architecture, highlighting the logical separation between the different functional layers, which allows for flexible implementation and potential extension to new data types or functionalities in the future.

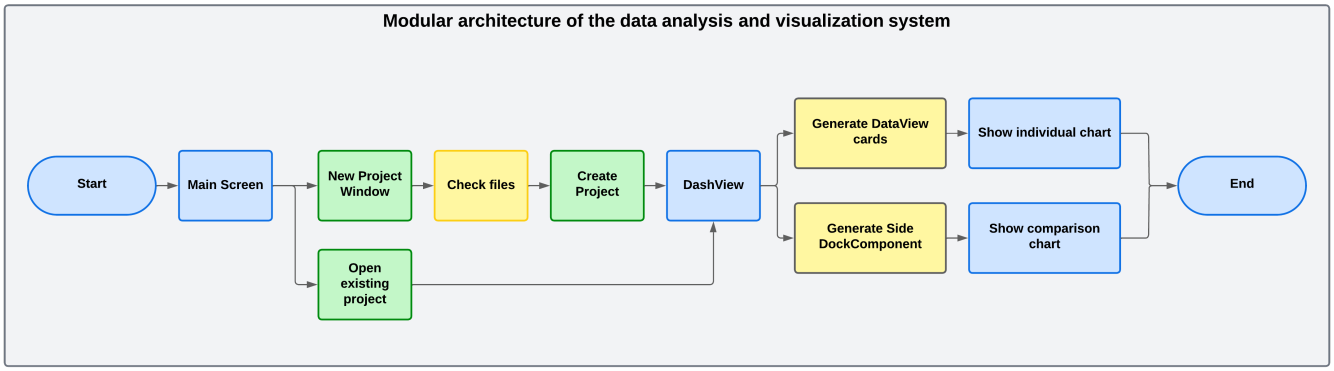

Fig. 3. General architecture of the proposed system for data analysis and visualization. The modular design clearly separates the input, processing, visualization, and control layers, enabling a flexible and scalable implementation.

C. Automated Statistical Analysis

The system implements an automated statistical analysis process on data collected from environmental sensors. This process begins with the gathering of multiple tabular data files, which are organized, renamed, and stored within a user-defined folder structure. The files are then merged into a single dataset to facilitate joint processing. Once integrated, the data are structured into different variables according to their nature (for example, temperature, pH, humidity, and others). Each variable undergoes a series of basic statistical calculations to extract information relevant to system monitoring, such as maximum, minimum, mean, median, mode, and variance. Additionally, a chronological analysis of the measurements is performed, allowing the identification of each variable's evolution over a specific day through its hourly distribution.

This statistical processing is intended to provide the end user with both a broad and detailed view of environmental variable behavior, facilitating the identification of patterns, anomalies, or critical conditions that may require attention within the monitored system.

D. Multivariate Predictive Modeling with Transformer Architecture

The predictive component of the developed tool is based on a deep-learning Transformer architecture, specifically designed to handle multivariate time-series sequences in the context of aquaponic systems. This model enables forecasting the behavior of five critical system variables—relative humidity, pH, TDS, pond temperature, and ambient temperature—using input windows of historical data. The proposed architecture is structured as a pure-encoder configuration in which the input consists of a fixed-length multivariate sequence of 96 steps, representing historical data for the five aforementioned variables. This input is first processed through a one-dimensional convolutional layer (Conv1D) to capture local temporal patterns. It is then passed through a normalization layer and a dense projection that adjust the internal dimensionality to 256 units, ensuring compatibility with the subsequent model blocks. To preserve the temporal position of each datum within the sequence, a positional-encoding mechanism using learnable embeddings is incorporated, allowing the model to distinguish the relative location of each time step without relying on recurrent layers. At the core of the model are four Transformer blocks, each comprising multi-head self-attention mechanisms followed by feedforward networks arranged sequentially. Within each block, self-attention operations enable the model to learn contextual relationships and long-range dependencies among variables across different time points. The output of each attention layer is merged with its input via residual connections and then normalized by a layer-normalization step to stabilize training. A distinctive feature of this architecture is a specialized branch dedicated to pH prediction. Recognizing that pH is especially sensitive and critical in aquaponic systems, the model splits the output of the final Transformer block into two pathways: (1) a general branch that applies global pooling and dense projections to forecast all variables jointly, and (2) a specialized branch that further processes the representation through a dedicated dense network for pH. The pH prediction from this branch then replaces the pH channel in the general output, thus providing a focused correction to improve accuracy.

The complete flow of the proposed model is illustrated in Figure 4, it details the main stages from the preprocessed input to the generation of multivariate predictions over a 24 hour horizon, showing the split into two output branches characteristic of the adopted Dual-Head approach. The model's final output is a multivariate sequence of 2,880 steps ahead equivalent to a 24 hour prediction horizon given the 30-second sampling frequency enabling simultaneous observation of all variables' anticipated behavior and thereby facilitating real time planning and decision making. The model was compiled using the Huber loss function suitable for environments with moderate outliers—and optimized with the Adam algorithm at an initial learning rate of 1×10^{-4}. Training was conducted for 300 epochs on Min–Max-normalized data, using a batch size of 64 samples and cross-validation on a hold-out test set.

The technical specifications of the system used to execute the model training are detailed in Table I.

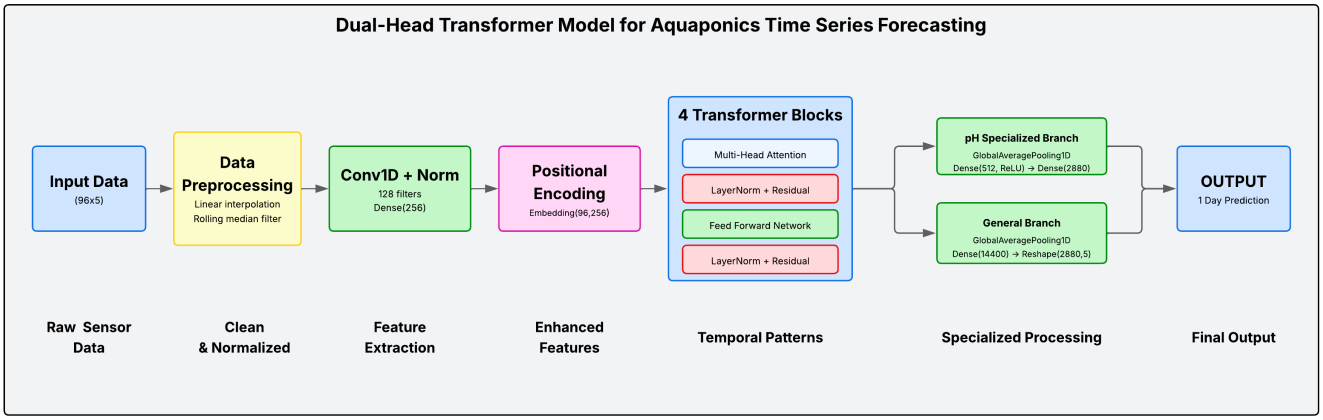

Fig. 4. General schematic of the architecture of the proposed Dual-Head Transformer model for multivariate prediction in aquaponic systems.

TABLE I
SPECIFICATIONS OF THE COMPUTING SYSTEM USED FOR MODEL TRAINING.

Component	Specification
Operating system	Windows 10 Pro (64 bits)
Processor (CPU)	AMD Ryzen 3 3100, 3.6 Ghz
RAM Memory	32 GB DDR4
Graphics processing unit	NVIDIA GTX 1650 4 GB GDDR6
Storage	SSD NVMe 120 GB
Development Framework	TensorFlow 2.15.0 (Keras API)
Programming language	Python 3.12
Training time	Approximately 28 hours.

III. RESULTS

A. Automatic dashboard and statistical analysis

The system generates an interactive dashboard designed to facilitate the analysis and interpretation of the variables monitored in the aquaponic system. This dashboard displays each variable through sliding cards that allow users to quickly view current values and recent behavior in a clear and accessible way. These cards function as individual panels that highlight the most relevant information for each parameter, optimizing the user experience and improving the speed at which changes or anomalies in the system can be detected.

In addition to the cards, the dashboard includes time-series charts for each monitored variable. These charts allow users to examine the evolution of the data over time, with the ability to explore different intervals and time scales, facilitating detailed analysis of trends, cyclical patterns, or unexpected fluctuations. This dynamic visualization is crucial for understanding how the variables respond to different environmental conditions or system adjustments.

Complementing the visual component, the lower section of each chart displays key descriptive statistics that quantitatively summarize the behavior of the variables. These metrics include the maximum and minimum values, which define the observed range; the mean and median, which provide central measures of the data; the mode, indicating the most frequent value; and the variance, which reflects the dispersion and stability of the variable under study. These indicators offer a solid foundation for data interpretation, enabling producers and technicians to make informed and timely decisions to optimize the operation and health of the aquaponic system.

The Figure 5 It illustrates a snapshot of the resulting dashboard, highlighting both the informative cards and the integrated descriptive statistical data.

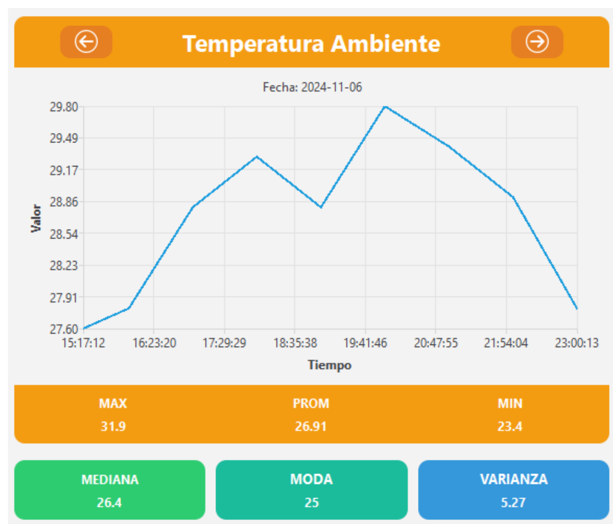

Fig. 5. Automatic dashboard with variable visualization and descriptive statistics.

B. Multivariate Prediction Tool

A deep learning model based on a specialized dual-output architecture was implemented for multivariate prediction of critical variables in an aquaponic system. This multivariate Transformer architecture incorporates a general branch for all variables and a specialized branch for pH, enhancing prediction accuracy for this particularly sensitive variable.

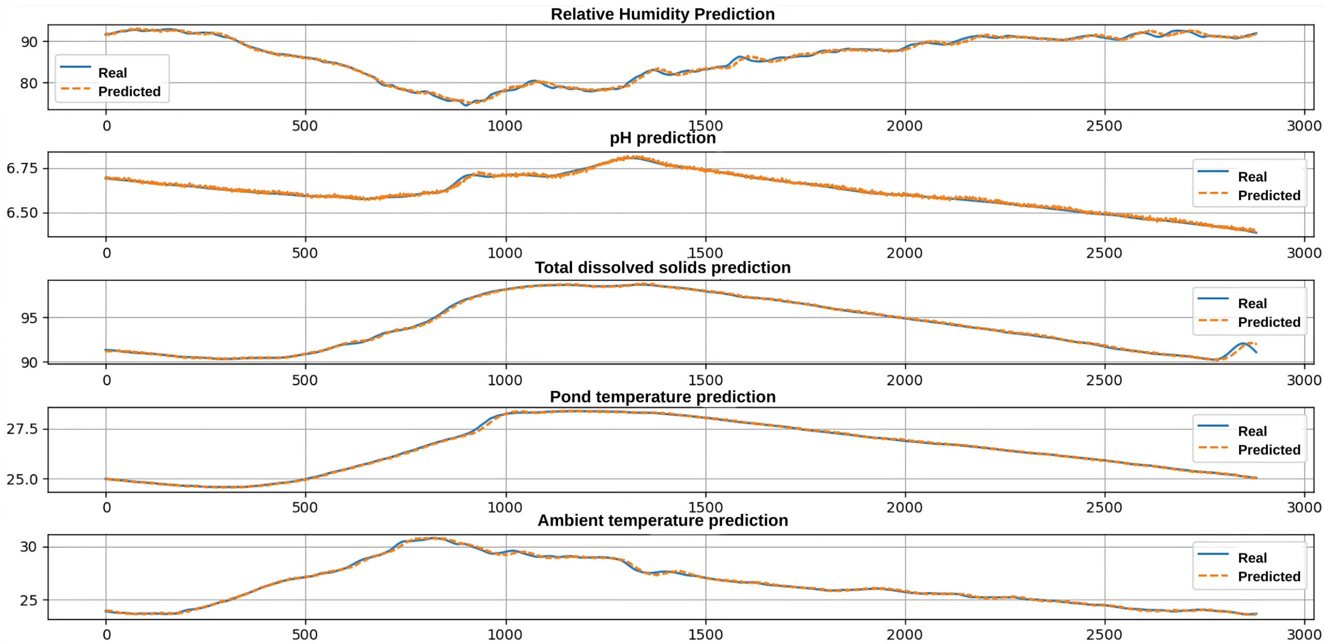

Fig. 6. Comparison between real and predicted values for a full day of operation (2,880 steps) across each of the five monitored variables.

The model takes as input time sequences of 96 steps and produces an output of 2,880 future steps, equivalent to a full day of prediction at a 30-second sampling frequency. The variables considered were: relative humidity, pH, TDS, pond temperature, and ambient temperature.

The training was conducted using a dataset collected in a real local aquaponic production environment, with 80 % of the samples for training and 20 % for validation and testing. The accuracy metrics obtained for the full prediction horizon are presented in Table II. The table includes the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2) for each of the evaluated variables.

TABLE II
PERFORMANCE METRICS OF THE PREDICTION MODEL FOR EACH VARIABLE.

Variable	MAE	RMSE	R^2
Humidity	1.0417	1.8875	0.8934
pH	0.1954	0.2345	0.8867
TDS	1.2739	1.7343	0.8435
Pond temperature	0.3899	0.5305	0.9307
Ambient temperature	0.7027	1.0153	0.9051

The results show a good fit of the model on the test set, with R^2 values above 0.84 for all variables, indicating that the model is able to explain a considerable proportion of the observed variance. Likewise, the low MAE and RMSE values relative to the typical operational range of each variable demonstrate that the predictions generated are sufficiently accurate to support monitoring, control, and decision-making tasks within the aquaponic system.

In the Figure 6 The predictions generated for an example from the test set are presented over a 24-hour horizon. Each subplot compares the real and predicted series for the five monitored variables, allowing observation of the model's fit and its ability to track the system's temporal dynamics. A close alignment between the real and predicted curves is evident, demonstrating the model's accuracy in forecasting system behavior. Relative humidity shows smooth daily variations that are correctly replicated by the prediction. In the case of pH, the prediction successfully captures small gradual changes and inflections, validating the effectiveness of the model's specialized branch. The TDS, pond temperature, and ambient temperature series exhibit well-defined continuous trends that the model follows precisely, even during moments of slope changes. These results demonstrate that the proposed architecture is capable of faithfully anticipating the multivariate evolution of the aquaponic environment.

IV. CONCLUSIONS

This research developed a comprehensive computational tool that combines automated statistical analysis, interactive data visualization, and a multivariate prediction model based on a specialized Transformer architecture with a dedicated branch to improve pH estimation accuracy. This solution enables continuous monitoring and in-depth interpretation of the dynamic behavior of critical variables such as relative humidity, pH, TDS, pond temperature, and ambient temperature, while also providing reliable 24 hour horizon forecasts. These capabilities not only strengthen proactive decision making by producers but also help prevent risks associated with system management, optimizing resource use and increasing operational efficiency of the cultivation. Validation of the system with real data obtained from a productive aquaponic environment with mixed crops and freshwater fish demonstrated the effectiveness of the proposed approach. This work represents not only a significant technical advance for the digitalization of aquaponic systems but also provides a solid foundation for its replicability, scalability, and future integration into broader agricultural settings, thereby promoting smart, sustainable agriculture aligned with the principles of Agriculture 4.0.

References

4 [1] T. Zou, A. Dawodu, E. Mangi, and A. Cheshmehzangi, "Exploring current trends, gaps & challenges in sustainable food systems studies: The need of developing urban food systems frameworks for sustainable cities," *Sustainability*, vol. 15, no. 13, p. 10248, 2023.

9 [2] C. S. Nair, R. Manoharan, D. Nishanth, R. Subramanian, E. Neumann, and A. Jaleel, "Recent advancements in aquaponics with special emphasis on its sustainability," *Journal of the World Aquaculture Society*, vol. 56, no. 1, p. e13116, 2025.

8 [3] L. Souto Cavalli, F. Blanco Marques, A. Watterson, and A. Ferretto da Rocha, "Aquaculture's role in latin america and caribbean and updated data production," *Aquaculture Research*, vol. 52, no. 9, pp. 4019–4025, 2021.

2 [4] R. d. Sousa, L. Bragança, M. V. da Silva, and R. S. Oliveira, "Challenges and solutions for sustainable food systems: The potential of home hydroponics," *Sustainability*, vol. 16, no. 2, p. 817, 2024.

10 [5] A. AlShrouf, "Hydroponics, aeroponic and aquaponic as compared with conventional farming," *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 27, no. 1, p.
12 247–255, Jan. 2017. [Online]. Available: https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/2543

6 [6] M. Krastanova, I. Sirakov, S. Ivanova-Kirilova, D. Yarkov, and P. Orozova, "Aquaponic systems: Biological and technological parameters," *Biotechnology & Biotechnological Equipment*, vol. 36, no. 1, pp. 305–316, 2022.

[7] B. Vásquez-Pineda and K. S. P. Hernandez, "Design of a monitoring
14 system for aquaponic crops using internet of things (iot)," in *2024 IEEE Central America and Panama Student Conference (CONESCAPAN)*. IEEE, 2024, pp. 1–6.

7 [8] M. Alselek, J. M. Alcaraz-Calero, J. Segura-Garcia, and Q. Wang, "Water iot monitoring system for aquaponics health and fishery applications," *Sensors*, vol. 22, no. 19, p. 7679, 2022.

5 [9] G. Y. Sriram and N. S. Shibu, "Design and implementation of automated aquaponic system with real-time remote monitoring," in *2021 Advanced Communication Technologies and Signal Processing (ACTS)*. IEEE, 2021, pp. 1–6.

3 [10] Y. Pan, J. Sun, H. Yu, G. Bai, Y. Ge, J. Luck, and T. Awada, "Trans-
3 forming agriculture with intelligent data management and insights," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 3489–3498.

[11] L. Bizon and M. Castro, "Aquaponics in brazil: review and survey on
13 waste management practices," *Revista AIDIS de Ingeniería y Ciencias Ambientales Investigación desarrollo y práctica*, vol. 16, pp. 518–534, 08 2023.