Lightweight Embedded Neural Network for Water Quality Classification (CONAMA 357/2005)

Eduardo Caldeira Vicente
Polytechnic School, University of Vale do Itajaí
Itajaí, Brazil
eduardo.c@edu.univali.br

Anita Maria da Rocha Fernandes
Polytechnic School, University of Vale do Itajaí
Itajaí, Brazil

Luis Augusto Silva
Department of Computer Science, Faculty of Science,
University of Salamanca
Salamanca, Spain

Wemerson D. Parreira
Polytechnic School, Faculty of Electrical Engineering,
Pontifical Catholic University of Campinas
Campinas, Brazil

Abstract

A compact neural-network model was developed to classify water samples according to quality levels established by CONAMA Resolution 357/2005. The study used a 2018 SEMASA-Itajaí dataset comprising 8,928 hourly measurements of pH, turbidity, apparent colour and local precipitation. After outlier filtering, weighted k-NN imputation and min–max scaling, the records were divided into a 70% training set and a 30% validation set. The final classifier, implemented in TensorFlow with two dense hidden layers and trained for 100 epochs with the Adam optimiser, achieved 98.7% accuracy, with misclassifications occurring only between neighbouring regulatory classes. Alongside the modelling, this work also proposes a conceptual design for a future data-visualisation layer intended for deployment in an embedded AIoT monitoring system.

Keywords

Artificial Intelligence, Neural Network, Water Monitoring, IoT

1 Introduction

Preserving water quality is a fundamental pillar of the United Nations 2030 Agenda, particularly within Goal 6, which focuses on universal access to safe drinking water and adequate sanitation. In 2022, more than 2 billion people were still without secure access to this essential resource, reinforcing the urgency for monitoring systems that are increasingly precise and low-cost [21]. Within this scenario, machine learning techniques (especially Artificial Neural Networks (ANNs)) have gained visibility due to their capacity to represent nonlinear interactions among multiple physicochemical indicators of water [2, 18].

Recent literature points out that combining ANNs with deep learning strategies and advanced preprocessing methods substantially improves the predictive performance of parameters such as turbidity, pH, conductivity, and chlorophyll-a [5, 16].

Simultaneously, the emergence of edge devices based on microcontrollers (such as the ESP32) and wireless communication technologies has strengthened Artificial Intelligence of Things (AIoT) architectures, a paradigm that integrates artificial intelligence with IoT infrastructure *to enable the acquisition, processing, and transmission of environmental data* in real time, promoting reductions in operational costs and enabling immediate corrective responses [14].

In Brazil, CONAMA Resolution No. 357/2005 defines the criteria for classifying water bodies and establishes emission limits for effluents [4]. Observing this regulation is crucial not only for environmental protection but also for safeguarding public health and ensuring the sustainability of economic sectors that rely on water resources. Studies such as Silva (2022) [20] emphasize how machine learning approaches can support the analysis of anthropogenic pressures on water quality, while applied research demonstrates the practical viability of intelligent monitoring systems in the field [2, 14].

Considering this context, the objective of this work is to design and validate a neural network capable of classifying water samples according to the criteria of CONAMA Resolution 357/2005, using the dataset provided by SEMASA-Itajaí. Beyond the modelling itself, the study also outlines a conceptual proposal for a future data-visualisation layer intended to support the deployment of an embedded AIoT monitoring device. Although this visual interface has not been implemented, it establishes the design principles for presenting real-time measurements, model predictions and regulatory classifications in a user-centric manner. The ultimate goal is to deliver an accessible, high-performance tool aligned with public policy guidelines, contributing to the sustainable management of water resources and advancing public health.

2 Development

This work is defined as applied research [17], whose objective is to develop a neural network capable of classifying surface-water samples according to the parameters established by CONAMA Resolution 357/2005. For this purpose, the study employed the dataset provided by the Municipal Water, Basic Sanitation, and Infrastructure Service (SEMASA) of Itajaí. A qualitative approach [13] was adopted to evaluate the accuracy of the model, supporting its later implementation on a microcontroller and the construction of an AIoT device.

Figure 1 summarizes the workflow, and each phase is described throughout this article. The process begins with obtaining the dataset and segmenting it based on the three available attributes. Subsequently, exploratory analysis and preprocessing are applied, including the imputation of missing data using the *K-Nearest Neighbors Imputer*. Afterwards, the rainfall index (retrieved through the

Computer On The Beach '17, Itajaí, Brazil

Open-Meteo API) is incorporated according to the sample collection dates. Once the dataset is complete, the neural network is trained, and the final step consists of validating the obtained results.

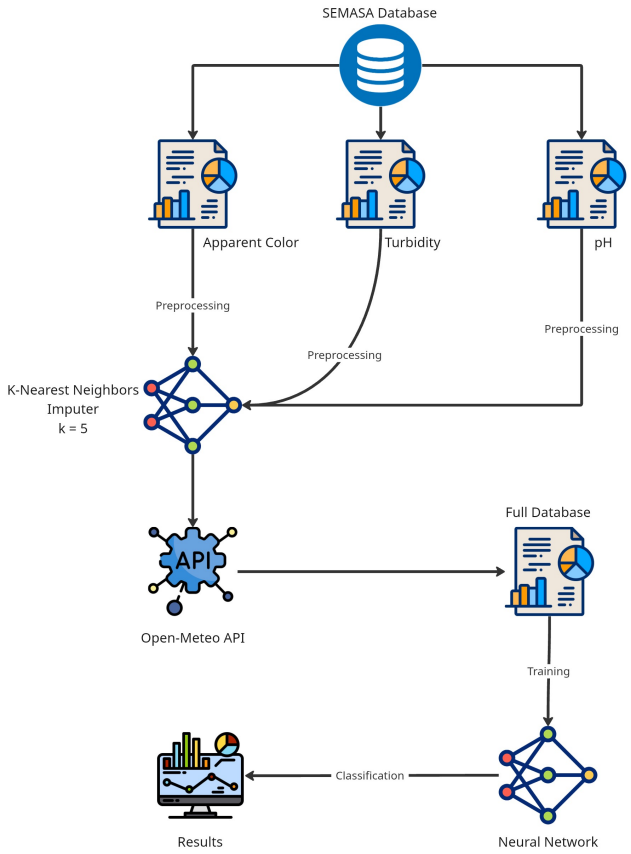

Figure 1: Process Overview diagram

To facilitate a clear understanding of the methodological procedures, the development of the study was structured into five stages, listed below:

- **Data Analysis, Standardization, and Formatting**: A detailed examination of the structure of the records was carried out, verifying distribution patterns, layout consistency, formatting conventions, punctuation, and naming standards. This stage included consolidating multiple schemas into a unified structure and defining consistent rules for data types, measurement units, and coding, ensuring compatibility with automated reading and processing by the artificial neural network.
- **Data Preprocessing**: Activities included identifying and handling missing entries, analyzing correlations among variables, computing descriptive statistics, and detecting or removing *outliers*, thus ensuring data integrity prior to training;
- **Data Imputation**: Missing information was filled using a machine-learning-based imputation technique, generating values coherent with the dataset's internal patterns. Additionally, a new feature representing the rainfall index for the collection date was incorporated;
- **Artificial Neural Network Development**: The chosen architecture was implemented in `Python`, with subsequent configuration of hyperparameters, and selection of the loss function and optimizer, followed by supervised training supported by cross-validation procedures;
- **Results Analysis**: Performance was assessed through metrics such as accuracy and by comparing the outcomes with an earlier version of the neural model.
- **Data Visualization**: A conceptual dashboard was outlined to guide the future visual presentation of measurements and model outputs. Although not implemented, this stage defines how real-time data, time-series behaviour and classification results could be displayed in an embedded AIoT interface, supporting interpretability and operational decision-making.

2.1 Database analysis, standardization and formatting

The dataset used in this study was supplied by SEMASA, the Municipal Water, Basic Sanitation, and Infrastructure Service of Itajaí. It contains records from the year 2018 collected at a monitoring station located in the São Roque neighborhood.

All variables present in the dataset were used for training the network, which are:

- pH measurement;
- Apparent Color measurement;
- Turbidity measurement;
- Rainfall Index (added later via API query).

The dataset provided by SEMASA was previously labeled into four categories (Class 1, Class 2, Class 3, and Class 4) based on the criteria established by CONAMA Resolution 357/2005. Theoretically, the resolution defines specific numerical intervals for these classes (e.g., pH limits usually bounded between 6.0 and 9.0, with varying maximum thresholds for turbidity and apparent color across the

four classes). However, instead of applying a deterministic, rule-based algorithm (which would merely replicate strict numerical thresholds), this study employs a neural network to implicitly learn the underlying patterns, non-linearities, and multi-variable interactions directly from the historical data. This approach allows the system to classify water quality more robustly, effectively handling real-world data variability and sensor noise.

Given the absence of a unified standard that would permit direct processing by the neural network, coupled with the dispersion of parameters across multiple files, it became necessary to restructure the dataset. The initial procedure consisted of merging all monthly spreadsheets into a single table, organized with fields for year, month, day, hour, and the respective value of each measured parameter.

In the sequence, every parameter and additional variable was incorporated into this unified structure, each arranged in its own column. This process generated a final dataset comprising 8,928 records. Entries with a value of zero were reclassified as missing data so that the subsequent algorithmic stages could handle them correctly.

2.2 Data Preprocessing

Data preprocessing was conducted using the R programming language and included: (i) analysis of missing data; (ii) calculation of the correlation matrix. The results of these steps are presented below.

2.2.1 Missing Data. During the preprocessing stage, one of the most relevant findings was the detection of missing values within the dataset. The variables turbidity and pH presented a relatively small proportion of absent entries (approximately 4%), while the apparent color variable exhibited a substantially higher rate, close to 50%, as shown in Figure 2. Because of this elevated percentage, the possibility of removing the apparent color variable and training the network without it was considered; however, the resulting models displayed instability and low accuracy. Consequently, the decision was made to impute the missing values for all three parameters (pH, apparent color, and turbidity) prior to initiating the training stage.

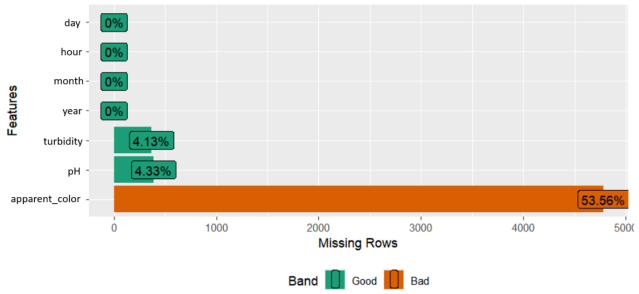

Figure 2: Missing data

2.2.2 Variable Correlation Matrix. To enhance the understanding of how the dataset variables interact, a correlation matrix analysis was conducted to assess the degree of influence among them.

The results showed that, beyond the expected strong correlations between repeated measurements of the same parameter (such as pH with pH), there is a notable moderate correlation (0.49) between apparent color and turbidity, as presented in Figure 3. This stronger association arises because variations in turbidity directly alter the visual coloration of the water.

A simple illustration of this phenomenon is observed when soil is introduced into a container filled with water and stirred: the mixture acquires a brownish tone, becomes turbid, and may even prevent visual transparency. The analysis also identifies a modest correlation between turbidity and apparent color with rainfall, suggesting that precipitation exerts a direct impact on these physicochemical attributes of the water and, consequently, on its overall quality.

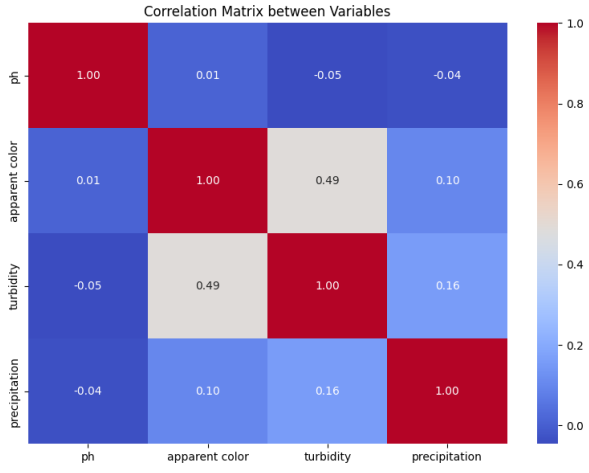

Figure 3: Variable Correlation Matrix

3 Data Imputation and Enrichment

3.1 Missing Values Diagnosis

The preliminary examination indicated the presence of gaps in the physicochemical parameters *pH*, apparent color, and turbidity, distributed unevenly across the dataset, as described in Section 2.2.1. This pattern is consistent with a *Missing At Random* (MAR) mechanism, which supports the application of imputation techniques based on similarity among samples [7].

Recent research highlights that weighted versions of the *k-Nearest Neighbors Imputer* (k-NNI) tend to outperform univariate approaches in scenarios with substantial multicollinearity, as they better preserve inter-attribute correlations [10, 12]. Moreover, hierarchical adaptations of k-NNI show stable and reliable behavior when preserving local relationships in environmental datasets, preventing the distortion of inter-attribute correlations [22].

In this study, the k-NNI method was employed with $k = 5$, using a normalized Euclidean distance metric and weights inversely proportional to distance. The algorithm was executed exclusively on numerical variables after standardization using the z-score procedure, following the workflow outlined in Figure 1.

3.2 Rainfall Index Integration

Additionally, the variable *daily precipitation* (mm), obtained through the *Open-Meteo Historical API* [23], was incorporated using the geographic coordinates of the São Roque neighborhood (26.911926°S, 48.719348°W), as illustrated in Figure 4. The literature identifies rainfall as an important explanatory factor for sudden variations in turbidity and apparent color [6].

To prevent excessive load on the external service, a local cache was implemented using a Python dictionary, allowing reuse of previously retrieved responses. This solution also addressed the slow request time of the API, which would have considerably increased the total processing time if every record required a new query. Because the API supplies daily(not hourly) precipitation data, the retrieved value remained constant for all twenty-four hourly records of each day. With the use of the cache, the process was reduced from potentially 24 requests per day to a single daily request.

Figure 4: SEMASA Sampling Location

The new column for precipitation was saved in the dataset file, providing a complete dataset without missing values and enriched with environmental context for the neural network modeling stage described in the following section.

3.3 Artificial Neural Network Development

The neural network was developed in Python using the TensorFlow framework [1]. The dataset was split into 70% for training and 30% for validation, a partition widely adopted in machine learning studies. The architecture comprises three fully connected layers: the first contains 32 neurons with ReLU activation, introducing nonlinearity and enabling the extraction of complex patterns [19]; the second also uses ReLU, with 16 neurons; and the final layer applies the Softmax function, responsible for converting raw output values into probability distributions suitable for multi-class classification tasks (specifically, CONAMA Classes 1 through 4) [3, 15]. This minimalist architecture results in only 756 trainable parameters, establishing an objectively lightweight profile suitable for memory-constrained edge devices.

During compilation, the Adam optimizer was selected, with a learning rate of 0.001. Adam integrates the advantages of Momentum (accelerating convergence by leveraging a moving average of past gradients) and RMSProp, which adaptively adjusts learning rates based on the moving average of squared gradients. This hybrid structure makes Adam particularly robust and efficient in deep learning applications involving large-scale data [9]. The loss function chosen was categorical cross-entropy, standard in multi-class problems where output predictions represent probability distributions. Cross-entropy measures the divergence between predicted and actual class distributions, driving the network toward improved predictions across training iterations [8, 11].

Training was carried out over 100 epochs, with mini-batches of 32 samples, using accuracy as the primary performance indicator. Model evaluation included overall accuracy, the confusion matrix comparing predicted and true classes, and the examination of accuracy and loss curves throughout the training process.

The original version of the network differed in both structure and configuration: precipitation was not included as an input attribute, and the hyperparameters were simpler—three dense layers with 16 neurons in the first hidden layer and 8 in the second, while the output layer remained the same. Adam was also used during compilation, although no explicit learning rate was defined. Training was initially executed for 50 epochs with batches of 16 samples. To enhance predictive performance, the precipitation variable was added, and an **empirical hyperparameter tuning** was conducted, manually testing combinations of neuron counts, learning rates, batch sizes, and training durations based on the model's convergence behavior.

3.4 Data Visualization

Although the neural network has been fully developed and validated, the Data Visualization stage remains at a conceptual level, serving as a proposal for the future implementation of the AIoT system. The visualization layer is designed to act as a bridge between the embedded model and the end user, translating predictions and measurements into clear graphical elements capable of supporting operational and environmental decision-making. To this end, a dashboard prototype (Figure 5) was created to represent the ideal organization of the information to be displayed by a real-time monitoring device.

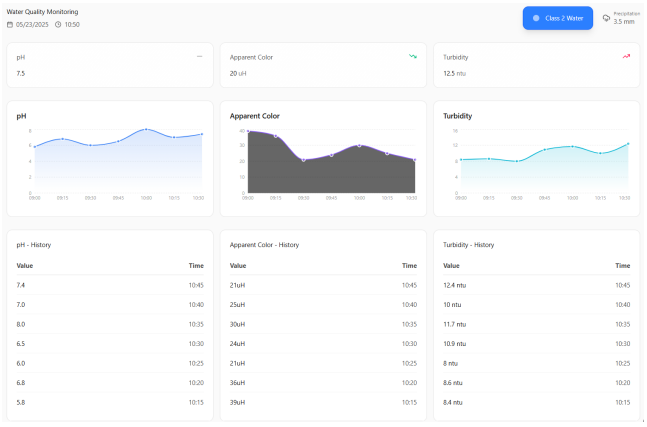

Figure 5: Data dashboard

The proposed concept includes three main blocks: instantaneous indicators of pH, turbidity, and apparent colour; time-series graphs illustrating the recent behaviour of these variables; and historical tables for detailed inspection. The structure follows the characteristics observed in the dataset and in the model's behaviour described in the previous sections. In addition to the three main parameters, the header displays essential information such as current date and time, precipitation level (obtained via API), and the focal point of the present research: the water-quality classification generated by the developed neural network.

4 Results and Discussions

This section describes the results obtained from the application of the techniques outlined in Section 2.

4.1 Artificial Neural Network Model

The model reached an accuracy of 98.7% during the validation phase, which used 30% of the dataset. The confusion matrix presented in Figure 6 shows that the few classification errors were concentrated in neighboring categories (one class above or below the true label) indicating coherent and stable behavior of the classifier. When compared to the initial version of the neural network, which obtained an accuracy of 98%, a considerably strong result, the refinements applied and the introduction of the precipitation variable produced an additional improvement of 0.7% in overall performance.

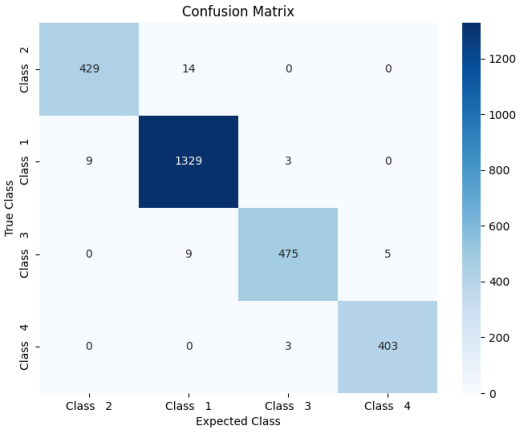

Figure 6: Confusion Matrix Predicted vs. Actual Classes

The confusion matrix also indicates that the misclassification errors are concentrated near the boundary thresholds of the classes, suggesting that these limits may require refinement. The use of *clustering* techniques to adjust the classification intervals could help mitigate such marginal discrepancies and, consequently, enhance the model's overall performance.

Another aspect of interest is the rapid convergence exhibited by the network, with accuracy stabilizing around the 40th epoch, as shown in Figure 7. This reflects the effectiveness of the adopted architecture, while also implying that further improvements (such as applying regularization strategies or exploring *data augmentation*) could be investigated to determine whether additional performance gains are achievable beyond the hyperparameter tuning already performed in this study.

Figure 7 displays the accuracy and loss curves throughout training and validation. For comparison, Figure 8 shows the performance of the first version of the model. The updated network demonstrates increased stability, with smoother and more consistent learning behavior, free of the oscillations observed in the earlier configuration. This improvement results from defining the Adam learning rate explicitly, including the new input variable, and refining the hyperparameters.

Finally, it is important to highlight that the model maintained stable predictive performance on the validation set even after the application of missing-data imputation. The low error rates achieved confirm that the preprocessing steps were successful and did not introduce relevant biases during training.

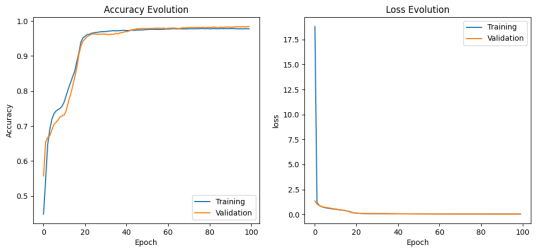

Figure 7: Accuracy and Loss Graphs of the New Artificial Neural Network

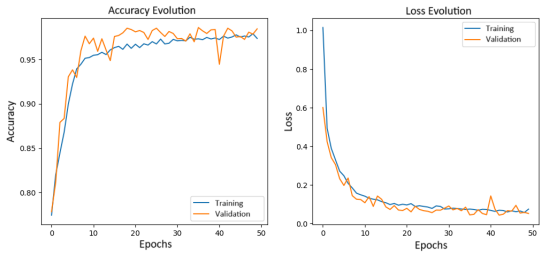

Figure 8: Accuracy and Loss Graphs of the Previous Artificial Neural Network

5 Conclusions

The study successfully developed and validated a lightweight neural network capable of classifying surface-water samples in accordance with the categories defined by CONAMA Resolution 357/2005, achieving an accuracy of 98.7%. The refinements introduced like explicit learning-rate definition, the inclusion of precipitation, expanded hidden layers, and systematic hyperparameter tuning resulted in a model with rapid convergence around the 40th epoch and significantly smoother accuracy and loss curves compared with the initial version. Misclassifications remained restricted to adjacent regulatory classes, indicating coherent and stable behaviour consistent with the expected transitions near class boundaries.

The methodological workflow, which encompassed restructuring fragmented spreadsheets, preprocessing, correlation analysis, missing-data diagnosis, weighted k-NN imputation, and dataset enrichment with daily rainfall, proved effective for organising a heterogeneous collection of physicochemical observations. The k-NN–based imputation preserved relevant correlations (especially the strong interaction between turbidity and apparent colour) allowing the network to maintain generalisation capacity even with nearly 50% missing entries in one of the core variables. These results demonstrate that the adopted pipeline is suitable for operational contexts in which historical monitoring datasets are incomplete or irregularly formatted.

Despite this strong performance, the study also reveals limitations intrinsic to the available dataset. Incorporating additional physicochemical attributes like dissolved oxygen or electrical conductivity may also strengthen the model's ability to capture subtle variations in water quality, provided that robust and synchronised datasets are available.

From an implementation standpoint, the compact architecture and low computational demand reinforce the feasibility of deploying the classifier on microcontroller-based AIoT devices. Embedded inference reduces latency, ensures operational autonomy in environments with intermittent connectivity, and lowers infrastructure requirements. In parallel, this work outlined a conceptual Data Visualization layer, defining how real-time measurements, temporal behaviour, and classification outputs could be presented in a user-centred dashboard. Although conceptual, this design establishes structural principles for a future interface to be integrated into an embedded monitoring system.

Furthermore, future studies should employ k-fold cross-validation and independent test sets to statistically ensure robust generalization capabilities, mitigating the limitations of the single 70/30 hold-out split used in this preliminary validation.

Future research should focus on real-time validation through physical sensors installed in the target environment, enabling the assessment of performance under operational noise, the detection of distribution shifts between historical and live datasets, and the establishment of periodic retraining cycles supported by continuous-learning mechanisms. Expanding the dataset to include seasonal variability, extreme weather events, and additional sampling points will be crucial for improving precision and ensuring broader applicability. Altogether, the study provides a technically sound foundation for intelligent water-quality monitoring, aligning machine learning, embedded systems, and regulatory compliance toward scalable and sustainable AIoT solutions.

Acknowledgments

The author gratefully acknowledges the support of the Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC). This research was carried out with the assistance of a FAPESC scholarship.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ Software available from tensorflow.org.

[2] Anderson Francisco de Sousa Almeida, Adonney Allan de Oliveira Veras, Bruno Merlin, Adam Santos, and Marcos Amaris. 2021. Predição temporal de parâmetros da qualidade da água usando redes neurais profundas. *Brazilian Journal of Development* 7, 11 (2021), 107662–107678. doi:10.34117/bjdv7n11-410

[3] Behnam Asadi and Hui Jiang. 2020. On Approximation Capabilities of ReLU Activation and Softmax Output Layer in Neural Networks. arXiv:2002.04060 [cs.LG] https://arxiv.org/abs/2002.04060

[4] Brasil. 2005. Resolução CONAMA nº 357, de 17 de março de 2005. https://conama.mma.gov.br/?option=com_sisconama&task=arquivo. download&id=450.

[5] G. Dharmarathne, A.M.S.R. Abekoon, M. Bogahawaththa, J. Alawatugoda, and D.P.P. Meddage. 2025. A Review of Machine Learning and Internet-of-Things on the Water Quality Assessment: Methods, Applications and Future Trends. *Results in Engineering* 26 (2025), 105182. doi:10.1016/j.rineng.2025.105182

[6] C. Varadharajan et al. 2022. Can Machine Learning Accelerate Process Understanding and Decision-Relevant Predictions of River Water Quality? *Hydrological Processes* 36, 12 (2022), e14565. doi:10.1002/hyp.14565

[7] P.J. Garcia-Laencina, J.L. Sancho-Gomez, and A.R. Figueiras-Vidal. 2010. Pattern Classification with Missing Data: A Review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 40, 4 (2010), 421–432. doi:10.1109/TSMCC.2010.2042428

[8] Elliott Gordon-Rodriguez, Gabriel Loaiza-Ganem, Geoff Pleiss, and John P. Cunningham. 2020. Uses and Abuses of the Cross-Entropy Loss: Case Studies in Modern Deep Learning. arXiv:2011.05231 [stat.ML] https://arxiv.org/abs/2011.05231

[9] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs.LG] https://arxiv.org/abs/1412.6980

[10] R. Kumar, S. Goyal, and M. Mittal. 2021. A Novel Weighted KNN-Based Approach for Missing Value Imputation in Multivariate Time Series Data. *IEEE Transactions on Knowledge and Data Engineering* (2021). doi:10.1109/TKDE.2021.3064772

[11] Anqi Mao, Mehryar Mohri, and Yutao Zhong. 2023. Cross-Entropy Loss Functions: Theoretical Analysis and Applications. arXiv:2304.07288 [cs.LG] https://arxiv.org/abs/2304.07288

[12] A. Mitra, A. Roy, and S. Khatun. 2021. A Hybrid KNN-Based Missing Data Imputation Method Using Enhanced Similarity Measure and Effective Feature Weighting. *IEEE Access* 9 (2021), 77157–77167. doi:10.1109/ACCESS.2021.3085552

[13] José Luis Neves. 1996. Pesquisa qualitativa: características, usos e possibilidades. *Caderno de pesquisas em administração* 1, 3 (1996), 1–5.

[14] N.S. Okomba, O.A. Esan, B.A. Omodunbi, S.A. Sobowale, F.D. Iyoaya, L.O. Nwobodo, and U.I. Nduanya. 2024. Development of Microcontroller Based Water Quality Monitoring and Water Level Control Device. *FUOYE Journal of Engineering and Technology* 9, 1 (2024), 43–48. doi:10.4314/fuoyejet.v9i1.7

[15] Anders Oland, Aayush Bansal, Roger B. Dannenberg, and Bhiksha Raj. 2017. Be Careful What You Backpropagate: A Case For Linear Output Activations Gradient Boosting. arXiv:1707.04199 [cs.LG] https://arxiv.org/abs/1707.04199

[16] C. Pacheco and N. Pereira. 2018. Deep Learning: Conceitos e Utilização nas Diversas Áreas do Conhecimento. *Revista Ada Lovelace* 2 (2018), 34–49.

[17] Cleber Cristiano Prodanov and Ernani Cesar De Freitas. 2013. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico* (2 ed.). Editora Feevale.

[18] Francisco de Assis Rocha, Maria de Jesus Delmiro e Souza Filho. 2021. Aplicação de redes neurais para a classificação e avaliação do grau de degradação da qualidade da água de reservatórios rurais no semiárido brasileiro. In *XXIV Simpósio Brasileiro de Recursos Hídricos.* https://anais.abrhidro.org.br/job.php?Job=12851 ISSN 2318-0358.

[19] Joseph Shenouda, Yamin Zhou, and Robert D. Nowak. 2024. ReLUs Are Sufficient for Learning Implicit Neural Representations. arXiv:2406.02529 [eess.IV] https://arxiv.org/abs/2406.02529

[20] C. V. Silva. 2022. *Uso de Redes Neurais Artificiais para Análise Multitemporal da Dinâmica do Uso e Cobertura da Terra em Bacias Hidrográficas*. Ph. D. Dissertation.

[21] UN-Water. 2023. Sustainable Development Goal 6 Synthesis Report on Water and Sanitation 2023: Blueprint for Acceleration. https://www.unwater.org/publications/sdg-6-synthesis-report-2023 Acessado em 19 de junho de 2025.

[22] Y. Yang, J. Darmont, F. Ravat, and O. Teste. 2022. Dimensional Data KNN-Based Imputation. In *Proc. Intl. Conf. on Advances in Databases and Information Systems (ADBIS)*. IEEE, 133–144. doi:10.1109/ADBIS.2022.00019

[23] Patrick Zippenfenig. 2024. *Open-Meteo.com Weather API (Version 1.0.0)*. Open-Meteo.com. doi:10.5281/zenodo.7970649 Historical Weather API documentation. Acesso em 20 jun. 2025.