Computers and Electronics in Agriculture 246 (2026) 111611

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Transformer-based artificial intelligence for forecasting energy demand in irrigation districts

Mariana Akemi Ikegawa Bernabé *, Rafael González Perea, Juan Antonio Rodríguez Díaz, Jorge García Morillo

Programa de Doctorado de Ingeniería Agraria, Alimentaria, Forestal y del Desarrollo Rural Sostenible. Universidad de Córdoba 14071 Córdoba, Spain

ARTICLE INFO

Keywords:
Transformer neural network
Time series forecasting
Deep learning
Energy demand in irrigation districts
Artificial intelligence

ABSTRACT

Efficient energy management in pressurized irrigation systems is essential to optimize water and energy use. This study presents a novel hybrid forecasting model based on Transformer Neural Networks (TNNs) integrated with fuzzy logic (FL) for mid-term hourly energy demand prediction in irrigation districts (ID). The model was applied to the Valle Inferior del Guadalquivir Irrigation District (VIGID), Spain, using real data from 2020 to 2023. A total of 26 potential input variables were initially considered, including climatic conditions, energy prices, crop distribution, and historical energy demand (*ED*). Through a rigorous statistical analysis combining FL and correlation matrices, 12 key variables were selected to capture the most relevant temporal and meteorological influences. A sequence-to-sequence architecture was implemented to model complex temporal dependencies in *ED* patterns. The model achieved high predictive performance, with an average coefficient of determination (R^2) of 99.62%, Root Mean Squared Error (RMSE) of 0.038 and Mean Absolute Error (MAE) of 32.36 kWh, demonstrating its ability to capture nonlinear behaviour and seasonal variability. The multi-head attention mechanism enhanced interpretability by dynamically weighting inputs according to context, while FL contributed to robustness under uncertainty. Results confirm the effectiveness of attention-based architectures for irrigation energy forecasting, supporting data-driven decision-making. Accurate *ED* prediction is key to optimizing resource allocation, reducing operational costs, and facilitating the integration of renewable energy sources into agricultural systems. This research underscores the potential of AI-driven tools to enhance energy efficiency and promote sustainability in modern irrigation practices.

1. Introduction

Sustainability concerns in irrigated agriculture, especially in semi-arid regions like Andalusia, Spain, have led to irrigation modernization aimed at saving water and improving efficiency (Rodríguez-Díaz et al., 2011). While replacing traditional channels with pressurized networks has reduced water use, it has greatly increased energy demand (*ED*), creating challenges for integrated water–energy management (Tarjuelo et al., 2015). Between 1950 and 2007, for example, irrigation water uses in Spain fell by 21%, while energy demand rose by 657% (Corominas, 2010). Modern on-demand systems provide farmers with continuous water access, but this flexibility introduces variability in energy use, making efficient resource management more challenging (Tarjuelo et al., 2015). Unlike urban or industrial energy systems, where demand patterns are generally more stable and centrally managed, energy demand in irrigated agriculture is strongly influenced by farmers' individual decisions, irrigation scheduling, and crop-specific requirements, introducing additional uncertainty and variability in consumption patterns. This intrinsic variability increases the exposure of irrigation systems to external market conditions, making energy costs particularly sensitive to electricity price fluctuations. Volatile electricity prices, driven by factors such as post-pandemic recovery and geopolitical crises, made energy a major cost factor, peaking in Spain in 2022 at 700 €/MWh before dropping back to pre-2021 levels in 2024 (OMI-Polo Espanol S.A. 2025).

For managers, it is important to know the expected *ED* in the coming days for aspects such as optimal tariff contracting. Moreover, in irrigated areas with solar power plants and surplus sales, this becomes even more critical because accurate forecasting is necessary to schedule energy fed into the electricity grid. In addition, given the increase in energy costs, irrigation districts (ID) seek to reduce expenses through photovoltaic solar energy, which makes it essential to have an accurate understanding

* Corresponding author.
E-mail address: p82ikikm@uco.es (M.A. Ikegawa Bernabé).

https://doi.org/10.1016/j.compag.2026.111611
Received 10 November 2025; Received in revised form 9 February 2026; Accepted 26 February 2026

of the energy needs of these communities in order to ensure proper management.

Artificial intelligence (AI) has demonstrated a remarkable capacity to address complex challenges in agriculture (Sheikh Khozani et al., 2022). In the context of pressurized irrigation systems, where farmers' decisions directly impact resource consumption, advanced models for forecasting *ED* and production emerge as key solutions to mitigate uncertainty and optimize operations. Hybrid AI models, which combine climatic, temporal, and operational variables, have proven particularly effective in generating reliable short- and medium-term predictions of water demand (González Perea et al., 2023, 2019a). Consequently, there is a growing interest in applying AI, Information and Communication Technologies and big data technologies in irrigated agriculture (Wolfert and Isakhanyan, 2022). These advanced approaches are crucial to improving the efficiency of resource management, especially the management of water-energy control in pressurized irrigation systems (Kamilaris and Prenafeta-Boldú, 2018) by forecasting water and energy demand according to sector-specific time series.

In 2017, Vaswani et al. (Vaswani et al., 2017) introduced a paradigm shift with the publication of their work on attention mechanisms in ANNs, which emphasized parallel attention-based architecture rather than sequential memory-based approaches. This breakthrough led to the development of Transformer architectures, initially designed for natural language processing (NLP) tasks, but which are increasingly adopted in time series forecasting due to their exceptional ability to capture complex relationships and long-term dependencies in sequential datasets. In contrast to memory-based mechanisms employed in LSTM and Gated Recurrent Unit (GRU) models, which rely on sequential information processing, attention mechanisms establish direct relationships between all elements of the sequence. This enables the model to focus selectively on the most relevant inputs, regardless of their temporal position. This architecture has been demonstrated to reduce training time and alleviate vanishing gradient issues, while also demonstrating superior scaling efficiency with large datasets.

In the context of irrigation and agriculture-specific forecasting, Abed et al. (Abed et al., 2022) employed a DL model for evaporation prediction in Malaysia, comparing its performance to other DL approaches such as Convolutional Neural Networks (CNNs) and LSTM models. González et al. (González Perea et al., 2024) developed a new model that combines a modified version of TNNs, FL, and GAs for middle-term irrigation water demand forecasting, enabling multi-step demand forecasting with a significant error reduction. Despite these advances, there is still a clear gap in the literature regarding the application of Transformer-based models for hourly energy demand forecasting in irrigated agriculture at a mid-term horizon (up to one week), which limits their practical use for operational energy management.

In this work, it has been developed a novel hybrid model for forecasting *ED*, which has been implemented in the Valle Inferior del Guadalquivir Irrigation District (VIGID), Spain. The model integrates a modified version of TNNs with FL to enhance hourly *ED* predictions at a mid-term horizon of up to one week ahead, using large historical data at the irrigation district level. By leveraging the Transformer architecture within a supervised learning framework for regression tasks, the model incorporates attention mechanisms to perform multi-step forecasting while effectively reducing prediction errors. This forecasting capability directly supports key operational decisions in irrigation districts, such as electricity tariff contracting and the coordination between grid-supplied energy and on-site photovoltaic generation, including the estimation of surplus energy to be fed into the grid. Its capacity to anticipate hourly *ED* up to seven days in advance represents a significant advancement, as most existing approaches in the literature focus on coarser temporal resolutions, such as daily or seasonal scales. This fine-grained temporal resolution enables more accurate daily planning, reduces billing penalties associated with forecasting errors, and improves overall water–energy management efficiency, contributing to greater operational efficiency and sustainability in ID. The remainder of this paper presents the study area and dataset, describes the methodological framework and model design, analyses the results and their implications for irrigation energy management, and concludes with a discussion on the relevance of the proposed approach and potential future research directions.

2. Related work

Machine learning (ML) techniques, such as artificial neural networks (ANNs), decision tree-based models, and metaheuristic optimization algorithms (e.g., genetic algorithms (GA) and particle swarm optimization), have significantly advanced the optimization of water and energy resource management (Behzadipour et al., 2023; Bernabé et al., 2025; González Perea et al., 2019b). For instance, González Perea et al. (González Perea et al., 2019a) introduced a hybrid methodology that combining ANNs, Bayesian framework and GA to model farmers' behaviour and predict irrigation depth at the farm level with short data sets. Recent studies have applied deep learning (DL) algorithms, particularly Long Short-Term Memory (LSTM) models, to model water consumption based on climatic data and scheduling inputs (Kavya et al., 2023), in building energy consumption (Deb et al., 2017) and in *ED* with real industrial data (Wang et al., 2020). These studies demonstrated that LSTM models often outperform traditional ML models in terms of reliability and accuracy. Furthermore, advanced hybrid models, such as LSTMHybrid, which integrate fuzzy logic (FL) and GA, have achieved high precision in demand forecasting (González Perea et al., 2023) and a Hybrid LSTM Neural Network for energy consumption forecasting of individual households (Yan et al., 2019). However, despite their advantages, LSTM models still present limitations such as difficulties in capturing very long-term dependencies and may suffer from overfitting when applied to complex and heterogeneous data.

Transformer-based architectures have recently emerged as a powerful alternative for time series forecasting, addressing limitations of recurrent networks by leveraging self-attention mechanisms to capture long-range dependencies and complex nonlinear patterns (Lin et al., 2022; Rigneault et al., 2020; Zhao et al., 2026). While Transformers have been increasingly applied to renewable energy production (Tian et al., 2022) and electricity load prediction (Saeed et al., 2025), energy demand forecasting presents distinct challenges. (Zhang and Wang, 2025). Unlike renewable generation, which depends primarily on climatic variables and system characteristics (Hussan et al., 2026), energy demand forecasting in irrigation presents unique challenges due to the impact of farmer behaviour, operational decisions, and crop schedules, variability not typical in urban or industrial systems. Variants such as Autoformer (Wu et al., 2022), Informer (Zhou et al., 2020), and ETSformer (Woo et al., 2022) improve accuracy and efficiency for long-sequence forecasting, with performance gains reported across energy, weather, and economic datasets.

Recent methodological advances focus on robustness and interpretability, including iterative estimation for stochastic systems (Liu and Ding, 2025), semantic analysis for contextual patterns (Zeng and Deng, 2025), multi-sensor data fusion (Iqbal et al., 2025), and PID control strategies for dynamic systems (Horstkötter and Zhu, 2025). These studies support the rationale for combining ML, LSTM, and Transformer-based architectures with hybrid methods integrating fuzzy logic and control theory for accurate irrigation energy demand forecasting.

3. Materials and methods

3.1. Case study

The developed model has been implemented and validated in the VIGID, located in Seville (Andalusia), southern Spain. The irrigated area covers 18,945 ha, which is divided into 3,074 farms spread across ten municipalities within the Guadalquivir basin, serving 2,300 farmers. The water distribution network is divided into nine sectors, each equipped with its own pumping station and regulation reservoir for

water storage dedicated to irrigation. Irrigation is on-demand, so water is continuously available to farmers. The nearest meteorological station is Villanueva del Río y Minas, located at UTM coordinates (x, y): (262609, 4164000). The average temperature in this area is 18.5°C, with an average maximum and minimum temperature of 26.1°C and 11.7°C, respectively. The annual average accumulated precipitation is 495 mm, while the annual average accumulated evapotranspiration is 1,295 mm, peaking in July with a daily average of 6.5 mm. In 2024, the main permanent crops in the VIGID were citrus (43.17%), followed by fruit trees (4.65%) and olive groves (4.21%). The primary annual crops included wheat (8.21%), cotton (7.83%), and potatoes (7.44%), together accounting for over 75% of the total cultivated area that year.

Each farm is fitted with a water meter that quantifies irrigation water consumption. These data are utilized to allocate the energy costs attributable to the operation of the irrigation water supply system. However, irrigation water use is not time-stamped; hence, its unit cost to farmer remains constant.

Since 2019, the VIGID has utilized photovoltaic energy from what was, at the time of this study, the largest solar facility among all Spanish irrigation systems (van de Loo et al., 2024). This plant has a capacity of 6 MWp and covers 14.37 ha. The solar panels are equipped with tracking mechanisms to maximize efficiency. Additionally, the photovoltaic plant is connected to the electrical grid, allowing any excess solar energy not used by the irrigation system to be sold to the national grid (Cayuela et al., 2024). A telemetry system monitors the solar energy generated, consumed, and exported to the grid. On average, slightly more than 50% of the energy used comes from the solar power plant and the remainder comes from the electrical grid. Each day, managers must predict how much energy will be needed, how much the solar plant will produce, and consequently, how much it will be fed into the electrical grid. Currently, this is done with low accuracy, resulting in billing penalties.

3.2. Problem approach

In this work, we focus on forecasting irrigation *ED*, in the middle-term (one week ahead) at the ID level, with an hourly time resolution, using large historical data. This is a time series forecasting problem, as current *ED* varies in a way that depends on previous *ED*, maintaining a temporal relationship within the *ED* sequence.

A well-known structure called sequence-to-sequence has been used in NLP tasks, as well as the encoder-decoder architecture for multi-step time series forecasting problems. This architecture is particularly effective for addressing such time series problems when the model's input and output lengths of the model follow a MIMO (Multiple Input, Multiple Output) structure, where the encoder processes the input sequence x and transforming it into a fixed-length context vector c, while the decoder applies transformations to this context to generate the final output sequence y (Allam and Mcewen, 2023).

The Transformer-based encoder-decoder architecture was chosen. The proposed TNN model was developed using an Intel Core i7-1195G7 CPU @ 2.90 GHz and 16 GB of RAM computer, built using TensorFlow 2.10.0 in Python 3.9.0.

3.3. Data collection

The model was trained and validated using hourly data spanning from January 2020 to December 2023. A total of 26 potential input variables ($PotI_X$) were identified, as detailed in Table 1 and categorized as follows: a) Climatic data: Maximum, minimum, and average temperature; b) Land use and crop data: Hectare distribution and crop percentage, considering only the five most representative crops; c) Energy price and tariff data: Electricity price and tariff periods; d) Temporal data: days of the year (DOY), weekdays, weekends, and months; e) Categorical data: Binary classification for day/night and precipitation occurrence; f) Historical *ED*: Hourly lagged values, specifically one hour before (t-1), two hours before (t-2), and three hours before (t-3).

Table 1
Potential model inputs variable ($PotI_X$).

Potential Inputs	Description	Temporal Resolution
$PotI_1$	Days of the year (DOY) (1–366)	Daily
$PotI_2$	Month (1–12)	Monthly
$PotI_3$	Weekday vs. Weekend (0–1)	Daily
$PotI_4$	Energy Tariff Period (1–6)	Hourly
$PotI_5$	Day vs. Night (0–1)	Hourly
$PotI_6$	Boolean Rainfall (0–1)	Daily
$PotI_7$	Maximum temperature (°C)	Daily
$PotI_8$	Average temperature (°C)	Daily
$PotI_9$	Minimum temperature (°C)	Daily
$PotI_{10}$	Maximum relative humidity (%)	Daily
$PotI_{11}$	Average relative humidity (%)	Daily
$PotI_{12}$	Minimum relative humidity (%)	Daily
$PotI_{13}$	Maximum wind speed (m/s)	Daily
$PotI_{14}$	Average wind speed (m/s)	Daily
$PotI_{15}$	Solar Radiation (MJ/m^2)	Daily
$PotI_{16}$	Rainfall (mm)	Daily
$PotI_{17}$	ET0 (mm)	Daily
$PotI_{18}$	Applied energy demand in the current hour (kWh)	15-minutes
$PotI_{19}$	Applied energy demand on the previous hour (kWh)	15-minutes
$PotI_{20}$	Applied energy demand on the two previous hours (kWh)	15-minutes
$PotI_{21}$	Applied energy demand the three previous hours (kWh)	15-minutes
$PotI_{22}$	Citrus area (%)	Yearly
$PotI_{23}$	Potatoes area (%)	Yearly
$PotI_{24}$	Cotton area (%)	Yearly
$PotI_{25}$	Wheat area (%)	Yearly
$PotI_{26}$	Sunflower (%)	Yearly

Following the time series analysis of the *ED*, it was determined that considering the three preceding hours as input variables was sufficient to capture the temporal dependencies necessary for model training.

The available datasets exhibited different temporal resolutions, Fig. 1. The available *ED* data had a 15-minute resolution, the available climate data had a daily temporal resolution, the energy price data were available monthly, and the crop data were annual. To ensure consistency and comparability, all data were resampled and unified to an hourly resolution for model analysis.

The climatic data were obtained from the Red de Información Agroclimática de Andalucía ("[dataset] Red de Información Agroclimática de Andalucía (RIA) | Instituto de Investigación y Formación Agraria y Pesquera (IFAPA)," n.d) of the Junta de Andalucía, Spain. Data on *ED*, tariff periods and cropping area were provided by the VIGID for this study.

3.4. Data preprocessing and subset generations

The raw *ED* data from the VIGID were first cleaned by removing negative values in the key variables, including grid import consumption and self-consumption from the photovoltaic plant. Total *ED* was computed as the sum of these components. The data were then resampled to hourly intervals by summing values within each hour. Any isolated quarter-hour missing values in the ED data were absorbed during the hourly aggregation. Temporal variables, such as daylight hours (*PotI*5), were calculated based on the geographical coordinates of the VIGID. After applying data filtering and cleaning processes, the final dataset comprised 35,065 measurements, corresponding to four irrigation seasons.

Given the time series nature of the problem and the wide range of scales among the model variables, both the input variables and the model targets were standardized to minimize the impact of unit and scale differences between features. Due to their variability, both datasets were standardized before being partitioned into training and validation sets. This standardization was performed using scikit-learn's

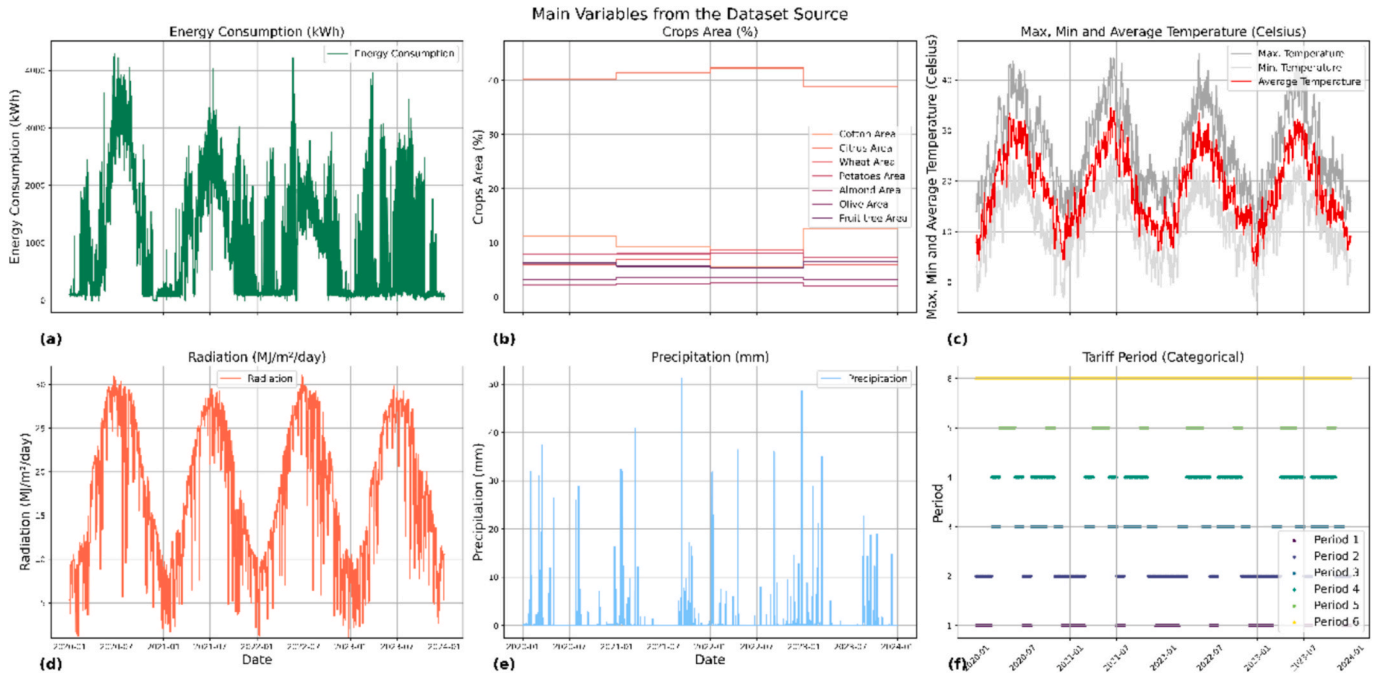

Fig. 1. Representation of main variables from the dataset .
source: (a) Energy Consumption; (b) Crops Area: cotton, citrus, wheat, potatoes, almonds, olive and fruit trees; (c) Temperature; (d) Solar Radiation; (e) Precipitation; (f) Tariff Period Classification

RobustScaler, which centres the data by subtracting the median and scales it according to the interquartile range (IQR), defined as the distance between the 25th and 75th percentiles. Mathematically, for each selected input variable (IV_x), where x represents the number of selected input variables to train the model, the transformation can be expressed in vector form as:

$$IV_x' = \frac{IV_x - median(IV)}{Q3(IV_x) - Q1(IV_x)} \quad (1)$$

where IV_x' is the vector of standardized values,$median(IV_x)$ is the median, and $Q1(IV_x)$ and $Q3(IV_x)$ are the 25th and 75th percentiles, respectively. The median and interquartile range are then stored to be applied to later data using the *transform* method.

The original dataset was randomly divided into two subsets with a split ratio of 85% for the training set and 15% for the validation set. To ensure good generalization, the split was performed using a Monte Carlo-based approach. Specifically, 100 random splits were generated, each using a different random seed, and the split that minimize distributional differences between the subsets was selected. This iterative procedure helps to reduce bias introduced by any particular partition and ensures that both the training and validation sets are statistically representative of the original dataset. In each split, inputs and targets were aligned with their corresponding timestamps to maintain temporal consistency, allowing for proper evaluation of the model on sequences without breaking chronological order. The Monte Carlo approach, combined with the Kolmogorov-Smirnov statistic (Corder and Foreman, 2009), guarantees that the distributions of key variables remain similar across the subsets, thereby improving model generalization and reducing the risk of overfitting.

3.5. Identification of forecasting model inputs (IV_x)

The features of one system are interrelated *ED* is influenced by multiple factors, such as weather conditions, temperature and humidity, energy price fluctuations (e.g., tariff periods), crop type, and past *ED*. Therefore, before incorporating them into the forecasting model, we must identify the significant inputs and reduce the dimension of the input space [38]. Selecting the most informative variables improves model performance and reduces computational complexity, making feature selection a critical step in developing accurate predictive models for *ED* (Lin et al., 1996).

In this study, FL curves were utilized to identify most representative input variables (IV_x) among the potential candidates ($PotI_X$) and a statistical analysis of the dataset was conducted to support the FL analysis. Specifically, the Pearson Correlation Coefficient (PCC) (Dutilleul et al., 2000) was applied to numerical variables, the Spearman Correlation Matrix (SCM) (Xiao et al., 2016) was used for discrete variables, and Point-Biserial Correlation (PBC) (Cureton, 1956) was employed for categorical data. It is important to compare these analysis with the FL analysis, as FL may present limitations when handling discrete and categorical variables. However, when variables exhibit more complex relationships, such dependencies may be lost in traditional statistical analyses. Moreover, an individual variable may show weak correlation, yet in combination with others, it can become highly predictive.

3.5.1. Fuzzy logic

FL curves are used to identify the IV_x, those that have the greatest influence on the model output, which in this case corresponds to *ED*. To achieve this, for each potential input variable, $PotI_X$, a fuzzy logic-based curve, denoted as c_X, is constructed in the $PotI_X - ED$ space, equation (2).

$$c_X(PotI_X) = \frac{\sum_{i=1}^{n} ED \times \mu_{X,i}(PotI_X)}{\sum_{i=1}^{n} \mu_{X,i}(PotI_X)} \quad (2)$$

where n is the total sample size, X is the total inputs variables and $\mu_{X,i}$ is the fuzzy membership function, defined as, equation (3).

$$\mu_{X,i}(PotI_X) = \exp\left(- \left(\frac{PotI_{X,i} - PotI_X}{b} \right)^2 \right) \quad (3)$$

where $PotI_{X,i}$ is the value of $PotI_X$ at point i in the $PotI_X - ED$ space and b is b 2 (Lin et al., 1996).

To quantitatively assess the relationship between input and output variables, we use the mean squared error (MSE), equation (4), between each fuzzy curve c_X and the original data. This measure allows us to

identify the most relevant input variables, IV_x. If the relationship is fully random, the fuzzy curve is flat and the MSE is large, which indicates a weak dependency. Conversely, a small MSE suggests a stronger correlation, implying that the corresponding input variable has a more significant influence (Lin et al., 1996).

$$MSE_{c_X} = \frac{1}{n}\sum_{i=1}^{n}\left(c_X(PotI_{X,i}) - \widehat{ED}\right)^2 \tag{4}$$

In a two-dimensional fuzzy curve, a fuzzy surface, s_{jk}, can be thought of. The fuzzy surface is defined in the following formula, equation (5).

$$s_{jk}(PotI_X, PotI_{jk}) = \frac{\sum_{i=1}^{n} ED \times \mu_{X,i}(PotI_X) \times \mu_{jk,i}(PotI_{jk})}{\sum_{i=1}^{n}\mu_{X,i}(PotI_X) \times \mu_{jk,i}(PotI_{jk})} \tag{5}$$

where $PotI_X$ and $PotI_{jk}$ are two potential inputs variables.

Hereafter, the MSE is computed for each fuzzy surface, denoted as $MSE_{s_{jk}}$, like equation (4). The input variables are then ranked in ascending order based on their MSE values. The $PotI_X$ is the most important input variable, identified by the fuzzy curves. Subsequently, the $PotI_j$ with the smallest $MSE_{s_{jk}}$ is considered the next most significant input, while the $PotI_k$ with the largest $MSE_{s_{jk}}$ is eliminated, (removing around 20% of the inputs by (Lin et al., 1996)). The fundamental principle behind using fuzzy surfaces to identify a set of independent input variables is to iteratively eliminate redundant or less relevant inputs at each step.

3.5.2. Correlation analysis

The correlation between two variables, denoted as $r_p(PotI_X, ED)$, measures the strength and direction of their relationship on a scale from -1 to $+1$. A coefficient of 0 indicates no correlation and 1 indicates stronger correlation, whereas values approaching ± 1 signify strong associations, positive or negative, respectively.

The PCC, where $PotI_X, ED$ are the vectors of the variables and n is the number of variables in each group, is suitable for measuring of strength and direction of only the the linear relationship between two continuous variables. The PCC is computed using the following formula, equation (6).

$$r_p(PotI_X, ED) = \frac{\sum_{i=1}^{n}(PotI_{Xi} - \overline{PotI_X})(ED_i - \overline{ED})}{\sqrt{\sum_{i=1}^{n}(PotI_{Xi} - PotI_X)^2}\cdot\sqrt{\sum_{i=1}^{n}(ED_i - \overline{ED})^2}} \tag{6}$$

where $\overline{PotI_X}$ and $\overline{ED}$ are the means of $PotI_X$ and ED, respectively. The numerator represents the covariance between $PotI_X$ and ED, and the denominator normalizes the covariance by the standard deviations of $PotI_X$ and ED.

The Spearman Correlation Matrix (SCM), denoted as $r_s(PotI_X, ED)$, is a nonparametric statistical measure that evaluates the strength and direction of a monotonic relationship between two ranked variables or between one ranked variable and one continuous variable. The SCM formula is equation (7).

$$r_s(PotI_X, ED) = 1 - \frac{6\sum_{i=1}^{n}(PotI_{Xi} - \overline{ED})^2}{n(n^2-1)} \tag{7}$$

Where numerator represents the difference between each pair of the ranked variables.

The PBC, denoted as $r_{pb}(PotI_X, ED)$, is a statistical measure used to analyse the relationship between a dichotomous variable and a continuous variable. In this work, we assume that the dichotomous variable (e.g. ED) takes two values 0 and 1. Under this assumption, the PBC is computed using the following formula, as presented in equation (8).

$$r_{pb}(PotI_X, ED) = \frac{(\overline{PotI_{X1}} - \overline{PotI_{X0}})}{\sqrt{\sum_{i=1}^{n}(PotI_{Xi} - \overline{PotI_X})^2}}\cdot\sqrt{\frac{n_1 n_0}{n^2}} \tag{8}$$

Where $\overline{PotI_{X1}}$ and $\overline{PotI_{X0}}$ are the mean on the continuous variable $PotI_X$ for all data points in$ED = 1$ and$ED = 0$, respectively. Further, n_1 and n_0 are the number of data points in$ED = 1$, n_0 and$ED = 0$, respectively, and n is the total sample size.

3.6. Transformer neural network architecture

The forecast model is an adaptation of the original Transformer architecture proposed by Vaswani et al. in (Vaswani et al., 2017). As shown in Fig. 2, the number of encoder and decoder layers (N_{enc} and N_{dec}) was reduced to 2, instead of the 6 layers used in the original model. The TNN follows a recursive encoder-decoder structure, incorporating multi-head attention mechanisms and positional encoding (PE) while utilizing teacher forcing during training. Teacher forcing is a training strategy whereby the model receives the ground-truth output from the previous step as input, rather than its own prediction (Williams and Zipser, 1989).

The encoder stack consists of an initial input sequence added to the PE layer and processed through a stack of N_{enc} identical encoder layers (Fig. 2). Similarly, the decoder stack is composed of N_{dec} identical layers. The input data maps the time series data to a vector of dimension d_{model} through a fully connected network. Each encoder layer consists of two sub-layers, as shown in Fig. 3.a. The first sub-layer is a multi-head Self-Attention mechanism (Fig. 3.c) and the second is a position-wise fully connected feed-forward network (FFN), Fig. 3.b. Both sub-layers are followed by a normalization layer to stabilize training and improve convergence.

The FFN within each layer serves as a dense neural network, analogous to the hidden layers in classical neural networks (Rumelhart et al., 1986). This sub-layer is composed of two linear activation function and applies a Rectified Linear Unit (ReLU) activation function, shown in Fig. 3.b. The ReLU was first introduced in the context of neural computation by (Hahnloser et al., 2000), who described "rectifying neurons" in a cortex-inspired model. Its efficacy in ANNs was subsequently popularised by (Nair and Hinton, 2010), who demonstrated that ReLU activation enhances the performance of Restricted Boltzmann

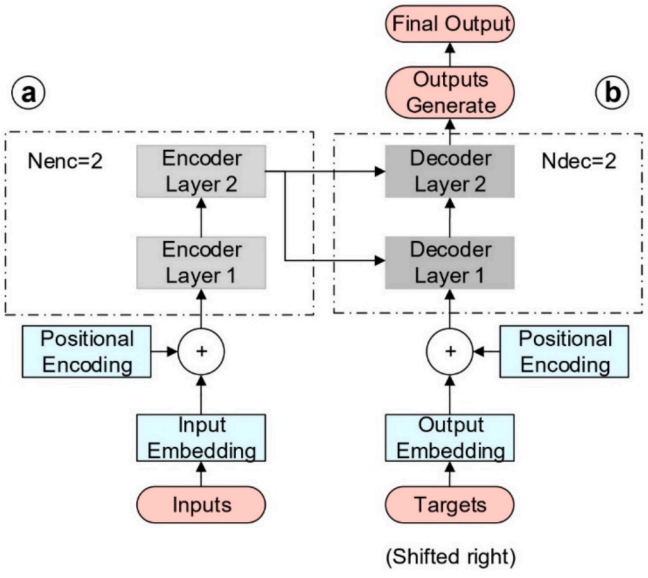

Fig. 2. The Transformer architecture with two encoder-decoder layers: (a) Encoder stack; (b) Decoder stack.

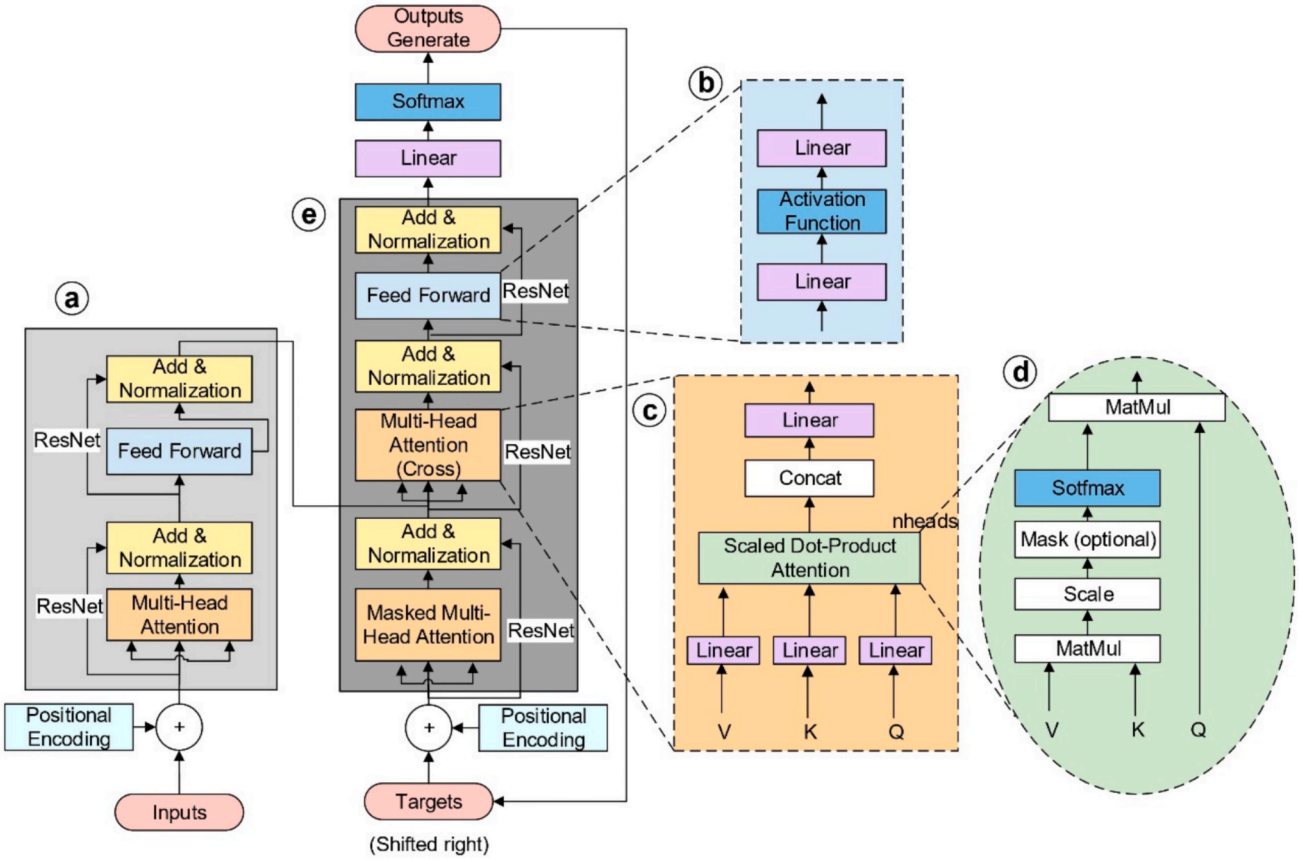

Fig. 3. The Transformer architecture and the attention mechanisms it uses: (a) Encoder layer; (b) Feed Forward Neural Network; (c) Multi-head attention; (d) Scaled dot-product attention; (e) Decoder layer.

Machines. The output of the FFN sub-layer is computed as follows equation (9):

$$FFN(x) = ReLU(W_1 + b_1)W_2 + b_2 \tag{9}$$

where W_1, W_2, b_1 and b_2 are the weight and bias parameters associated with the two hidden layers. The ReLU activation function, defined as equation (10):

$$f(x) = x^+ = \max(0, x) \tag{10}$$

ensures non-linearity by allowing only positive values to pass through while setting negative inputs to zero. This mechanism enhances the model's ability to capture complex patterns within the data.

In the decoder, an additional third sub-layer is introduced, dedicated to multi-head attention over the encoder's output, termed Cross-Attention, and shown in Fig. 3.e. Similar to the encoder, residual connections are applied around each sub-layer, followed by layer normalization. Additionally, the decoder employs a Masked Multi-Head Attention mechanism, termed Casual-Attention, which ensures that predictions at position i are conditioned only on known outputs from preceding positions. This is achieved through causal masking, combined with the one-step offset of the output embeddings, preventing the model from accessing future tokens during training.

The primary advantage of this architecture lies in its ability to leverage attention mechanisms to capture global dependencies between input and output sequences. Unlike traditional sequential models, such as LSTM or Autoregressive Integrated Moving Average (ARIMA), which process data step by step, the TNN enables parallel processing, significantly improving efficiency. However, since time series data inherently relies on temporal dependencies, the model integrates PE to preserve sequential information. This PE is implemented using sine and cosine functions, as described below.

3.6.1. Positional encoding for energy demand time series forecasting

Transformer-based models do not inherently preserve sequential information since they do not rely on recurrent connections or convolutional structures. Instead, they process input data in parallel, which allows for efficient computations but results in the loss of positional information. To compensate for this, PE is introduced to provide the model with an explicit understanding of the temporal order of input sequences.

PE is typically implemented as a stack of sine and cosine functions with varying frequencies (Vaswani et al., 2017). This approach ensures that each position within the sequence is uniquely represented while maintaining a continuous relationship between adjacent positions. The mathematical formulation of the sinusoidal PE is given by:

$$\begin{aligned} PE_{(pos,2i)} &= \sin\left(pos/1000^{2i/d_{model}}\right) \\ PE_{(pos,2i+1)} &= \cos\left(pos/1000^{2i/d_{model}}\right) \end{aligned} \tag{11}$$

Where pos represents the position in the sequence, i is the dimension index, and d_{model} is the embedding dimension. This formulation ensures that the encoding values follow a geometric progression, allowing the model to infer relative positions between tokens effectively.

The interleaving of sine and cosine functions creates a smooth encoding space where similar timestamps, such as consecutive days or months, remain close in the representation space. This is particularly relevant in time series forecasting, where it is essential to maintain the temporal relationship between observations. Furthermore, in multistep forecasting problems, additional temporal information, such as the hour of the day, day of the week, and month of the year, can be encoded using sine and cosine transformations to ensure periodic continuity.

By incorporating PE, the Transformer model gains awareness of the temporal structure of the input sequence, allowing the self-attention mechanism to consider the relative positions of tokens effectively. The use of sinusoidal encoding offers the advantage of extrapolating to sequence lengths beyond those encountered during training, thereby enhancing the model's generalization capabilities.

3.6.2. *Attention mechanisms*

Attention mechanisms are a fundamental component of TNNs, allowing the model to dynamically focus on relevant parts of the input sequence at different stages of processing. These mechanisms are used at three different levels, shown in Fig. 3.a and Fig. 3.e: Self-Attention in the encoder, where the encoder attends to different positions within the input time series; Masked Casual-Attention in the decoder, where the decoder attends to different positions within the target time series; and Cross-Attention, where the decoder layer attends to the encoder's output representations to generate predictions.

Following (Vaswani et al., 2017) original nomenclature, the attention mechanism relies on three vectors for each input element: query (Q), key (K), and value (V). These vectors are computed by multiplying the input with learned weight matrices Wq, Wk, and Wv. The core idea is that each value is weighted by a function of the query and its

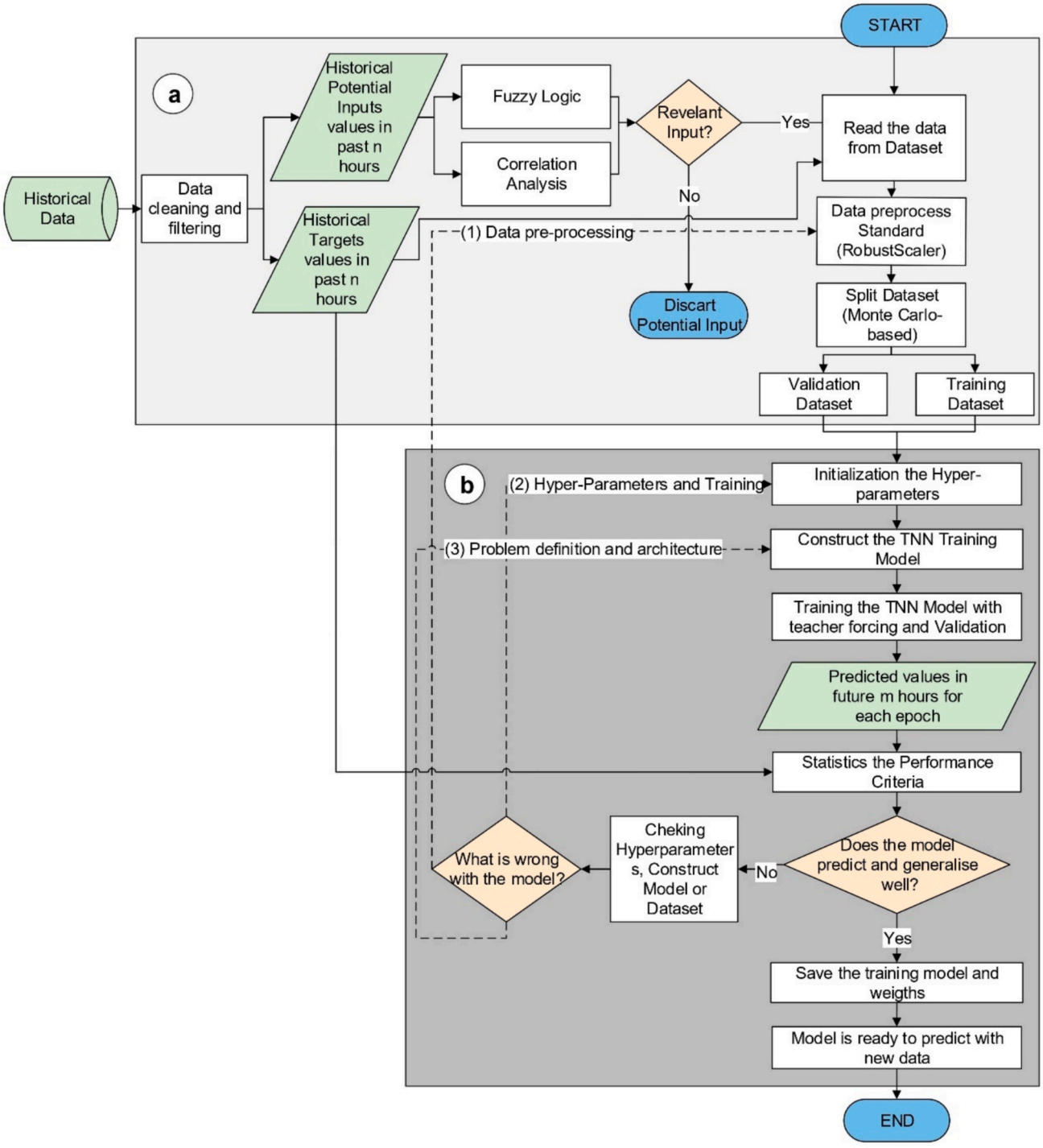

Fig. 4. Model Framework Flowchart: (a) Identification of the Forecasting Model Inputs; (b) Training and Validation Model.

corresponding key. The resulting weighted sum determines the output representation.

Mathematically, the attention mechanism is defined by the scaled dot-product attention, as shown in equation (12). This equation computes the attention scores by first obtaining the dot product of the Q and the K, which captures their similarity. The result is then scaled to prevent excessively large values, which could lead to unstable gradients.

$$\text{Attention}(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{12}$$

Where d_k is the dimension of the K.

The SoftMax function, defined in equation (13), normalizes the scores so that they are all positive and sum to 1, converting them into a probability distribution that determines the relative importance of each value V. The final output is a weighted sum of the values, where higher attention scores increase the contribution of specific values.

$$SoftMax = \frac{e^{zi}}{\sum_{j=1}^{m} e^{zj}} \tag{13}$$

for $i = 1, \cdots, m$ and $z = (z_i, \cdots, z_m) \in \mathbb{R}^m$.

To enhance the model's learning capacity, Transformers employ multi-head attention, equation (14), which allows the model to attend to different representation subspaces simultaneously. Multiple attention heads are used, each with its own set of learned projections. Each head computes attention independently, and their outputs are concatenated and transformed by a linear layer with learnable weight W^O.

$$\text{MultiHead}(Q,K,V) = Concat(head_1, \cdots, head_h)W^O \tag{14}$$

Where each attention head is defined as $head_h = Attention\left(QW_h{}^Q, KW_h{}^K, VW_h{}^V\right)$, where $QW_h{}^Q, KW_h{}^K$ and $VW_h{}^V$ denote distinct parameter matrices learned for each $.head_h$

The matrix W^O projects the concatenated outputs of all attention heads back to the model dimension through a final linear transformation. This linear layer integrates the information captured by the different heads, enabling the model to combine multiple attention perspectives into a coherent contextual representation.

3.7. Model framework

The model framework is displayed in the flowchart in Fig. 4. The dataset was divided into training and validation subsets following a Monte Carlo-based approach and standardized with the *RobustScaler*, as detailed in Section 3.3. Historical data of the model's potential inputs were analysed through FL analysis, and the least influential features were discarded, as described in Section 3.4. These procedures are consistent with step (a) of the proposed framework (Fig. 4), after which the data were used for model training and validation.

To improve training efficiency and model convergence, teacher forcing was employed during the training process. Specifically, the model was initially trained using a teacher forcing ratio of 1.0 (i.e., 100% of the ground truth output at each time step was fed as input to the decoder). This ratio was progressively decreased linearly with each epoch until it reached zero, allowing the model to gradually rely more on its own previous predictions. Formally, the teacher forcing ratio at epoch e out of a total of E epochs was calculated as follows equation (15):

$$TfRatio = \max\left(0, 1 - \frac{e}{E}\right) \tag{15}$$

During each training step, the application of teacher forcing was stochastic, controlled by this ratio. A random number r $U(0,1)$ was drawn, and teacher forcing was applied only if $r < TfRatio$. This probabilistic mechanism enabled a smooth transition from guided training towards autonomous sequence generation. Such an approach stabilizes early training, prevents error accumulation, and enhances the model's ability to generalize during inference.

The forecasting model was then constructed following the structure outlined in Section 3.5, and the hyperparameters were initialized. The model hyperparameters control both the architecture and the training process. The architectural hyperparameters considered include N_{enc}, N_{dec}, d_{model} $head_h$, the dimension of the FNN layer (d_{ff}), *dropout rate*, and the activation function of the FNN layer; while the training hyperparameters comprise *train batch size*, *optimizer*, *learning rate*, *loss function*, E, and *early stopping patience*. The model was subsequently trained and validated using the respective datasets, and its performance was evaluated based on predefined metrics. After generating hourly forecasts for a 7-day time horizon (168 h), an inverse transformation of standardization was applied to the predicted values, and the performance forecasting model was evaluated against the target data. Finally, the model's ability to generalize to unseen data was assessed. If the model failed to produce satisfactory predictions during the training phase, a systematic error analysis was conducted to identify potential issues, which could include: (1) Data preprocessing errors (e.g., an unrepresentative dataset split); (2) Inappropriate hyperparameter initialization or suboptimal training process; (3) Issues related to problem definition or model architecture.

In addition, an early stopping mechanism was devised to enhance the generalizability of the forecasting model. This mechanism was designed to monitor the loss function during the training process, thereby preventing the overfitting of the forecasting model. Specifically, the model tracked the validation loss $\left(L_{val}^{(e)}\right)$ value across successive epochs e, comparing it to the best recorded value (L_{best}). If a lower validation loss was detected $\left(L_{val}^{(e)} < L_{best}\right)$, the best value was updated, and the stagnation counter was reset. Otherwise, the stagnation counter S increased. The training process was terminated if no improvement was observed over ten consecutive epochs $(S \geq patience)$, preventing overfitting and reducing unnecessary computational costs.

To enhance model generalization and stability during training, a dropout was applied. This regularization technique randomly deactivates a fraction of neurons during each training step, preventing the model from becoming overly reliant on specific features and reducing the risk of overfitting. Additionally, the Adam the Adaptive Moment Estimation (ADAM) optimizer was used to ensure smooth and stable convergence. ADAM is an optimization algorithm that combines momentum and adaptive learning rates to improve gradient descent efficiency in training deep learning models. ADAM combines the benefits of momentum and adaptive learning rates, enabling more efficient gradient descent in deep learning models. Rather than using a fixed learning rate, a custom dynamic schedule was implemented, based on the formula proposed in (Vaswani et al., 2017). This schedule increases the learning rate linearly for a set number of warm-up steps, followed by a decay proportional to the inverse square root of the training step. Formally, the learning rate at each step is computed as follows equation (16):

$$LearningRate(step) = d_{model}^{-0.5} \cdot \min\left(step^{-0.5}, step \cdot warmup_{steps}{}^{-1.5}\right) \tag{16}$$

This approach allows for rapid learning in the initial training phase while ensuring stability and finer updates in later stages. It is particularly well-suited for TNN architectures, where maintaining an appropriate learning rate schedule is crucial for convergence and overall performance.

3.8. Performance evaluation

To assess the performance of the proposed hourly forecasting model based on the TNN algorithm, metric functions are employed in the evaluation process to quantify its accuracy. In this study, the MSE, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2) were considered as evaluation metrics. The

mathematical definitions of these evaluation criteria are given as follows, equation (17), 18, 19, 20).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ED_i - \widehat{ED_i})^2} \quad (17)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(ED_i - \widehat{\mathrm{ED}_i})^2 \quad (18)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|ED_i - \widehat{ED_i}| \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(ED_i - \widehat{ED_i})^2}{\sum_{i=1}^{n}(ED_i - \overline{ED_i})^2} \quad (20)$$

Where ED_i represents the observed value, $\widehat{ED_i}$ denotes the predicted output value, $\overline{ED_i}$ represents the average of the observed values and n is the total number of samples.

The R^2 indicates the proportion of variance in the target variable explained by the model, with values closer to 1 denoting better performance. The RMSE provides insight into the model's prediction error magnitude, giving more weight to larger errors, while the MSE quantifies the average of the squared differences between predicted and target values, making it useful for penalizing significant deviations. The MAE, on the other hand, measures the average absolute difference between predicted and observed values, offering a more interpretable error measure in the original scale of the target variable.

4. Results and discussion

4.1. Input variables for the forecasting model

4.1.1. Energy demand and irrigation context

Since energy is primarily consumed for water pumping, in pressurized irrigation networks, *ED* is intrinsically linked to irrigation water consumption, making it a key factor in irrigation operating costs. Irrigation seasons in VIGID runs from early May to mid-October, regulated by precipitation and the amount of water accumulated in the basins. This relationship shows a clear trend, which presents the corresponding irrigation season periods throughout the studied years, as well as the water restrictions experienced during this time. This figure highlights the periods of highest and lowest *ED*. In 2020, the maximum water allocation for the irrigation season was 4,500 m^3/ha, 2,700 m^3/ha in 2021, 1,000 m^3/ha in 2022, whereas in 2023, CHG approved an allocation of 700 m^3/ha for high-water-demand crops, representing a nearly 90% reduction compared to the concession allocation ("Comunidad de Regantes del Valle Inferior del Guadalquivir," n.d).

In 2020, the average maximum *ED* was approximately 3,500 kWhIn subsequent years, partly due to the annual water allocations authorized by the CHG for VIGID, influenced by water availability and the climatic conditions of each year. Specifically, the maximum *ED* decreased to 2,500 kWh in 20211,600 kWh in 2022, and 800 kWh in 2023. This monotonic reduction highlights the dominant role of institutional water management policies over purely meteorological drivers in shaping long-term ED dynamics. While previous works on irrigation water demand forecasting report interannual variability driven mainly by climate (e.g., temperature, radiation, and ETo) (González Perea et al., 2024, 2023), the present case demonstrates how regulatory constraints can override climatic signals, leading to structurally different energy consumption regimes across years.

Based on 35,064 measurements collected during the study period from 2020 to 2023, *ED* was from no demand to 4,299 kWh, with an average of approximately 818 kWh. Seventy-five percent of *ED* was less than or equal to 1,502 kWh. Fig. 5 presents the distribution of *ED* through statistical representations. Specifically, the relative frequency histogram (Fig. 5.a) reveals a highly skewed distribution, with a significant concentration of values at lower consumption levels and a long tail extending toward higher values, suggesting the presence of consumers with atypically high energy demands. The cumulative curve in Fig. 5.b confirms this trend, showing a rapid increase in the initial segments followed by a gradual stabilization. Therefore, it is necessary to understand and analyse the high temporal variability of the target of this work, *ED* in an ID. This analysis helps identifying where the model has most difficulties in the predictions. In this case, the range with higher values will be more challenging due to the limited amount of available data compared to the lower *ED* segments.

4.1.2. Statistical analysis of potential input variables using FL and correlation methods

In order to improve the performance of the forecasting model and to refine its potential input variables, FL and statistical analyses were performed, as described in Section 3.4. FL analysis was conducted to evaluate potential input variables for the model. The most frequent input variables were PotI1, PotI19, PotI20, and PotI21, with a 100% frequency, followed by high-frequency variables: PotI15 and PotI7. Additionally, variables with a frequency between 40% and 50% were identified, including PotI9, PotI22, PotI4, and PotI8, while the remaining variables had a frequency below 20%.

To cross-validate the results obtained in the previous analysis, correlation matrices were applied. Specifically, the PCC was used for numerical variables, and the SCM was applied to discrete variables. The numerical, discrete and categorical variables are those listed in Table 2.

The most influential numerical input variables were $PotI_{19}$, $PotI_{20}$,

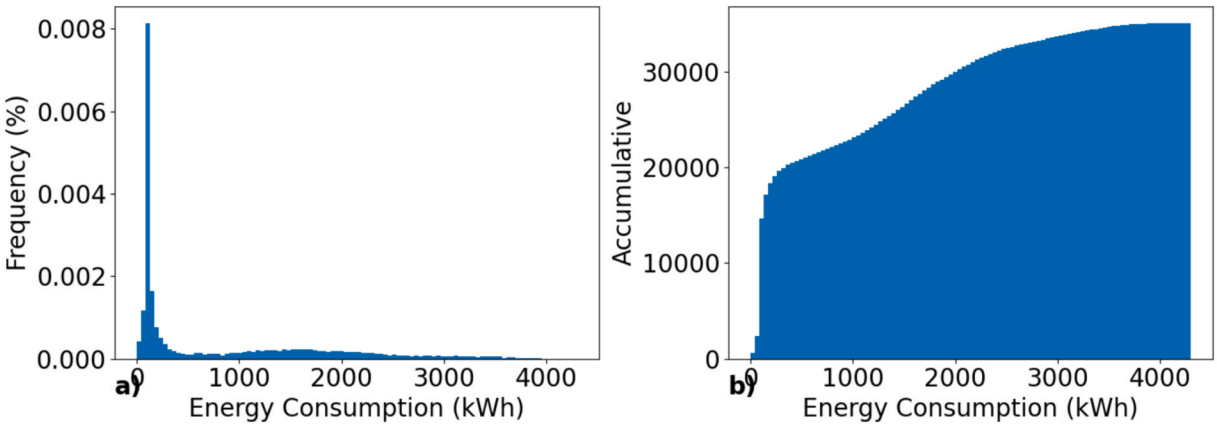

Fig. 5. Distribution of Energy Consumption (kWh) through statistical representations: (a) Relative frequency histogram; (b) Cumulative curve of energy consumption.

Table 2
Numerical, discrete and categorical variables.

Variables	Potential input variables ($PotI_X$)
Numerical	$PotI_7$, $PotI_8$, $PotI_9$, $PotI_{10}$, $PotI_{11}$, $PotI_{12}$, $PotI_{13}$, $PotI_{14}$, $PotI_{15}$, $PotI_{16}$, $PotI_{17}$, $PotI_{18}$, $PotI_{19}$, $PotI_{20}$, $PotI_{21}$
Discrete	$PotI_1$, $PotI_2$, $PotI_4$, $PotI_{22}$, $PotI_{23}$, $PotI_{24}$, $PotI_{25}$, $PotI_{26}$
Categorical	$PotI_3$, $PotI_5$, $PotI_6$

$PotI_{21}$, $PotI_7$, $PotI_9$, $PotI_8$, $PotI_{15}$, and $PotI_{17}$. The remaining numerical variables did not significantly impact *ED*.

In the SCM, none of the considered discrete variables ($PotI_1$, $PotI_2$, $PotI_4$, $PotI_{22}$, $PotI_{23}$, and $PotI_{26}$) showed a significant correlation with *ED*. However, $PotI_{22}$, which exhibited high influence in the FL analysis, was discarded in the SCM. This outcome is more consistent, as it is known that, in the specific case of VIGID, farmers do not adjust irrigation practices according to energy tariff periods, since they are charged a flat rate for the entire irrigation season. This behaviour contrasts with other energy demand forecasting contexts, such as building energy consumption, where tariff structures and user behaviour often introduce strong temporal dependencies, as reported in recent hybrid Transformer–Liquid Neural Network models (Antonesi et al., 2025). The absence of tariff-driven behaviour in VIGID reinforces the importance of context-aware variable selection and highlights the risk of blindly transferring assumptions from urban or residential energy forecasting to irrigation districts.

On the other hand, regarding the discrete variable $PotI_1$, although it did not show significant relevance in the SCM, it represents the DOY, which makes it a relevant input variable for *ED*. This is because *ED* is directly linked to irrigation seasons, which follow an annual temporal pattern.

Additionally, the relationship between potential categorical variables and EC_y was analysed (Fig. 6). In the PBC, was observed that $PotI_5$ and $PotI_6$ (Fig. 6.a and Fig. 6.c), which exhibit considerable variation between their categories, have an influence on *ED*. This result is expected, as *ED* is linked to irrigation water use, which is affected by both precipitation and daylight hours. Conversely, the variable $PotI_3$ did not show significant variation between weekdays and weekends, indicating that it does not influence *ED*. This finding further confirms that farmers' *ED* is independent of the tariff periods. This is logical since the water meters used are analog and therefore cannot record the time of consumption; as a result, farmers pay the same amount for each kWh regardless of when irrigation occurs.

4.1.3. Final input selection

After a comprehensive analysis of all results obtained from the different analysis approaches, the following 12 variables were selected as model input variables: $PotI_1$, $PotI_5$, $PotI_6$, $PotI_7$, $PotI_8$, $PotI_9$, $PotI_{15}$, $PotI_{17}$, $PotI_{18}$, $PotI_{19}$, $PotI_{20}$, and $PotI_{21}$. The remaining 14 candidate input variables were discarded. Table 3 presents the most influential input variables identified in each analysis and highlights the final selected input variables.

4.2. Forecasting model

4.2.1. Dataset, model configuration, and training setup

The VIGID dataset was used to develop the irrigation *ED* forecasting model. To ensure the reliability and transparency of the results, all performance and statistical metrics presented in this section were derived from the validation dataset (subgroup of VIGID dataset).

The model architecture, considering the limitations in computational capacity, consisted of 2 encoder layers and 2 decoder layers, with an input transformation space d_{model} of 256 dimensions. The model architecture also included 4 attention heads and an d_{ff} value of 1024. To reduce the risk of overfitting, a dropout rate of 10% was implemented, and the ADAM optimiser was utilised in conjunction with the learning rate schedule, which increases the learning rate linearly during a warm-up phase. Hyperparameter tuning included testing warm-up steps in the range of 2,000–8,000, a batch size of 96 constrained by computational resources, and early stopping with patience values between 5 and 20, ultimately selecting a patience of 10 epochs to balance convergence and overfitting prevention. Other parameters were set based on commonly used grid search values, ensuring robust and reproducible training of the forecasting model, shown in Table 4. The total number of epochs (E) used in the training of the model was 350.

4.2.2. Predictive performance and model behaviour

Direct experimental comparisons with LSTM and GRU models were not performed in this study, as previous works in a similar irrigation district, first applying an LSTM model (González Perea et al., 2023) and subsequently comparing it with a Transformer-based model (González Perea et al., 2024) demonstrated the clear superiority of Transformer architectures for multi-step water demand forecasting. Given the strong coupling between water and energy consumption in irrigation systems, these results can be reasonably extrapolated, justifying the focus on the

Table 3
Most Influential Input Variables: By analysis and selected final variables (IV_x).

Analysis	Most influential input variables
Fuzzy Logic Curves (FL)	$PotI_1$, $PotI_{19}$, $PotI_{20}$, $PotI_{21}$ (1.0); $PotI_{15}$ (0.8); $PotI_7$ (0.6); $PotI_9$ (0.5); $PotI_4$, $PotI_8$, $PotI_{22}$ (0.45);
Pearson Correlation (PCC)	$PotI_{19}$ (0.96); $PotI_{20}$ (0.93); $PotI_{21}$ (0.9); $PotI_{17}$ (0.63); $PotI_{15}$ (0.6); $PotI_8$ (0.56); $PotI_7$ (0.55); $PotI_9$ (0.49);
Spearman Correlation (SCM)	–
Point-Biserial Correlation (PBC)	$PotI_6$ (0.24); $PotI_5$(0.191);
Input variables selected (IV_x)	$PotI_1$, $PotI_5$, $PotI_6$, $PotI_7$, $PotI_8$, $PotI_9$, $PotI_{15}$, $PotI_{17}$, $PotI_{18}$, $PotI_{19}$, $PotI_{20}$, $PotI_{21}$

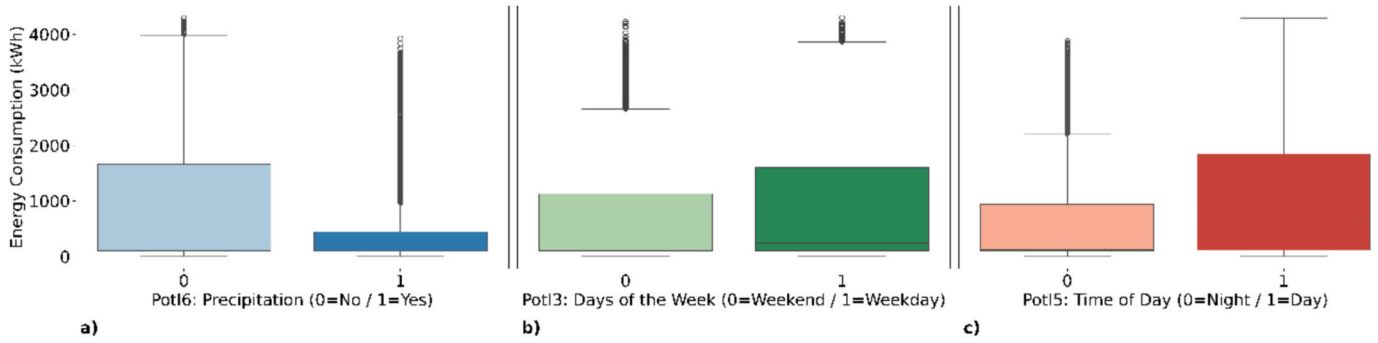

Fig. 6. Statistical analysis by point-biserial correlation for categorical variables: (a) $PotI_6$: Precipitation Boolean variables; (b) $PotI_3$: Days of the week (weekdays vs. weekends); (c) $PotI_5$: Daylight hours and night hours.

Table 4
Values of the hyperparameters selected for the forecasting model.

Variables	Values
N_{enc} and N_{dec}	2
d_{model}	256
$head_h$	4
d_{ff}	1024
Dropout rate	10%
Activation function of FNN	ReLU
Train Batch Size	96
Optimizer	ADAM
Learning rate	Custom Schedule by (Vaswani et al., 2017)
Loss Function	MSE
E	350
Early Stopping Patience	10
Encoder layer parameters	1,580,032
Decoder layer parameters	2,107,392
Dense layer parameters	257

Transformer-based model in the present study.

The encoder layer contained 1,580,032 parameters, while the decoder layer had 2,107,392. The final dense layer included 257 parameters, leading to a total of 3,687,681 parameters in the forecasting model. The model demonstrated excellent predictive performance, achieving an average R^2 of 99.62%, indicating that it explains nearly all variability in *ED*. The RMSE of 0.038 and the MSE of 0.0017 on scaled data reflect low prediction errors, with the RMSE emphasising larger deviations. The MAE was found to be 0.0246 on scaled data and 32.36 kWh when descaled, thus providing a clear estimate of the average absolute error in practical units. These results confirm the model's strong ability to accurately capture complex consumption patterns and its relevance for effective energy management in irrigation systems. This performance clearly exceeds that reported in earlier hybrid LSTM-based models for irrigation districts, where representativeness values ranged between 94% and 99% but with average prediction errors close to 20% (González Perea et al., 2023). It is also comparable to, and in some cases surpasses, the accuracy reported in recent Transformer-based irrigation water demand models, which achieved standard prediction errors around 2% using fewer parameters but focused exclusively on water demand rather than energy demand (González Perea et al., 2024).

Fig. 7 illustrates the correlation between the observed irrigation *ED* (actual values) and the predicted values at all forecasted time steps (t + 1 to t + 168) for the entire validation time series. The scatter plot (Fig. 7.a) demonstrates a strong linear relationship, with most of the points closely aligned with the ideal reference line (red dashed line), which represents a perfect prediction scenario (i.e., predicted = actual). This alignment suggests that the forecasting model achieves high predictive accuracy and generalizes well across the validation dataset. The residual analysis (Fig. 7.b) shows that most prediction errors are concentrated close to zero, with a small positive mean bias of 23.18 kWh, suggesting a slight tendency toward overestimation. The residual distribution exhibits a near-symmetric shape with low skewness (−0.16), while the majority of errors fall within the interquartile range from − 6.13 kWh (Q1) to 50.19 kWh (Q3). Extreme residuals are relatively infrequent, with approximately 90% of the errors lying between − 25.48 kWh and 112.81 kWh. Although the Shapiro–Wilk test indicates a deviation from normality ($p < 0.001$), the sharp peak around zero and the limited dispersion demonstrate a high concentration of small errors. Overall, these results indicate that the model provides robust and consistent forecasts across the validation period, with rare large deviations and stable performance over multiple prediction horizons.

A slight deviation is observed in the upper range of *ED* values, particularly around 4,000 kWh. This deviation is attributed to the lower density of hourly irrigation energy records in this range, which constrains the model's predictive performance due to the limited availability of historical data. These values correspond to peak consumption periods occurring only during the most intensive phases of the irrigation season. They are not considered outliers, as extreme values were previously processed. Instead, they represent less frequent but expected fluctuations in hourly *ED*, often driven by external factors such as high temperatures, strong cold winds, or a high concentration of irrigated crops.

4.2.3. *Practical and economic implications for irrigation management*

From a management perspective, a tool that predicts demand with such a low and consistent error level over time is highly valuable for decision-making. For example, knowing the *ED* in advance, and thus the water demand, is useful for planning reservoir filling, optimizing and maintaining pumping stations, and forecasting power requirements for energy tariff contracting. Additionally, VIGID also operates a large solar plant that sells surplus energy to the grid making reliable forecasts essential to anticipate the amount of energy to be injected each day and avoid financial penalties.

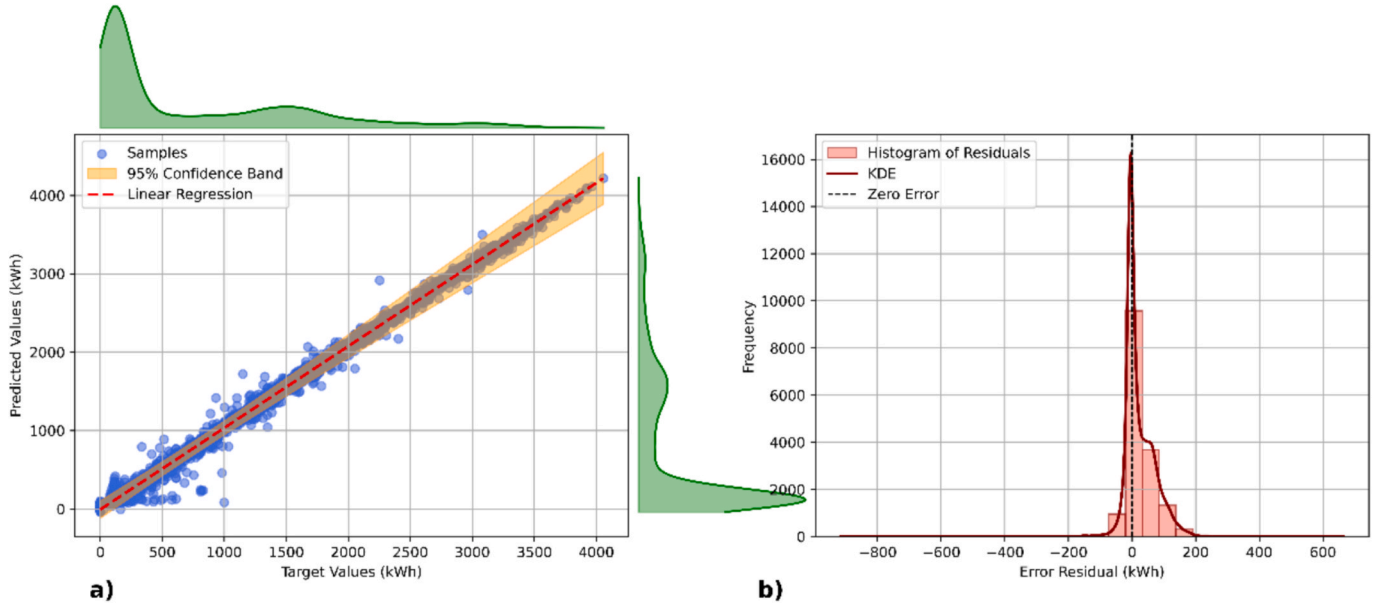

Fig. 7. Target and predicted irrigation energy demand of the complete validation dataset: (a) scatterplots; (b) histogram of residuals.

Beyond its operational advantages, mid-term *ED* forecasting carries direct economic implications for ID operating under demand-based schemes. Although the electricity cost is not transferred to farmers, the CHG is financially responsible for deviations between the forecasted and the actual energy injected into the grid. These deviations represent penalties that increase when the predicted photovoltaic generation or irrigation demand diverges from reality. By providing reliable seven-day-ahead forecasts, the proposed model enables managers to better align expected consumption with the photovoltaic energy available for sale, thereby reducing deviation costs and improving the economic return from renewable generation. This integration of predictive analytics into daily operations supports a more resilient and economically efficient management strategy, where energy and water planning are jointly optimized.

The methodology, based on attention mechanisms, provides additional insight into the interrelationships among the various factors influencing the actual water and *ED* in an ID. Farmers often make irrigation decisions based on socio-cultural factors rather than theoretical crop water requirements, so understanding these dynamics is crucial. The predictive capabilities of the proposed TNN model also have important implications for renewable energy integration. By forecasting hourly ED up to seven days in advance, managers can align irrigation schedules with photovoltaic energy availability, minimizing penalties from grid injection deviations. The attention-based mechanisms reveal which climatic and operational variables drive consumption at different times, enabling informed adjustments in pumping and energy allocation. Consequently, the model not only supports more efficient water and energy management but also lays the groundwork for integrating forecasting models into decision-support systems and optimization algorithms capable of jointly managing water and renewable energy resources across diverse operational contexts, contributing to both economic and environmental sustainability.

4.3. Multi-head attention mechanisms

The TNN architecture stands out due to its incorporation of attention mechanisms, which constitute one of its main strengths. These mechanisms enable the model to process all input variables simultaneously, assigning different weights to each one based on its relevance to the prediction task. Unlike memory-based architectures such as LSTMHybrid, which suffer from memory leakage effects as the forecasting horizon increases, the attention-based formulation used here allows the model to maintain stable performance across the entire seven-day horizon. This behaviour is consistent with findings reported in both irrigation-focused (González Perea et al., 2024) and wind-energy forecasting studies (Nascimento et al., 2023), where Transformers exhibited superior stability for medium-term predictions compared to recurrent models.

In addition, the use of multi-head attention enables the model to learn different patterns or aspects of the data in parallel, thereby enhancing its generalization capabilities. In the context of time series modelling, such as *ED* forecasting, these mechanisms are particularly valuable, as they allow the model to dynamically focus on the most influential factors at each time step, adapting to seasonal variations, hourly cycles, or even isolated events.

4.3.1. Analysis of cross-attention under real operating conditions

To analyse how these mechanisms operate under real-world conditions, beyond the training process, Figs. 8 to 11 are presented. These figures show hourly *ED* forecasts (from $t+1$ to $t+168$), along with the corresponding cross-attention maps for several time intervals randomly selected from the validation dataset. These maps provide insight into which input variables the model focuses on at each moment, offering an additional interpretability tool for understanding the system's behaviour. Four representative periods were selected, considering different stages of the irrigation season, sunlight availability, and the distinction between weekdays and weekends. These periods are: a) 23rd February 2022, 20:00: prior to the irrigation season, during nocturnal hours on a weekday (Fig. 8); b) 10th June 2020, 05:00: the timing of this occurrence corresponded to the period of maximum irrigation activity, which occurred during daylight hours on a weekday (Fig. 9); c) 30th July 2022, at 04:00: concurrently with the irrigation season, but during nocturnal hours on a weekend day (Fig. 10); and d) 29th November 2022, at 14:00: subsequent to the irrigation season, during daylight hours on a weekday (Fig. 11).

4.3.2. Seasonal and temporal attention patterns

During the irrigation season (Figs. 8 and 9), the model assigns greater attention to meteorological variables such as $PotI_6$, $PotI_7$, $PotI_8$, $PotI_9$, $PotI_{15}$, and $PotI_{17}$, which provide critical climate-related information (e.g., temperature, radiation, rainfall). Specifically, heads 1 and 2 tend to focus on $PotI_6$ and $PotI_7$ during weekdays, while heads 3 and 4 pay more attention to $PotI_8$, $PotI_9$, $PotI_{15}$, and $PotI_{17}$, especially on weekends, when solar radiation and evapotranspiration (ETo) become more

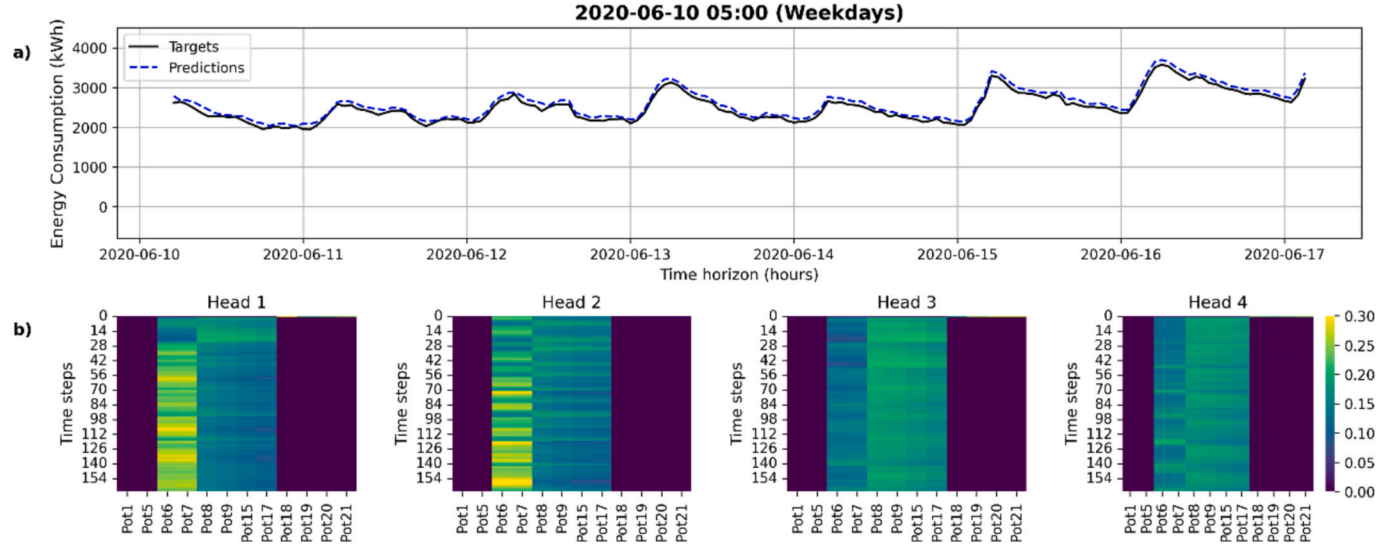

Fig. 8. Representation of the energy demand predictions and attention mechanisms for the selected model by 10-06-2020 at 05:00 (Weekdays): (a) target and predicted values of energy demand in kWh; (b) four attention heads of the cross-attention mechanism (encoder–decoder).

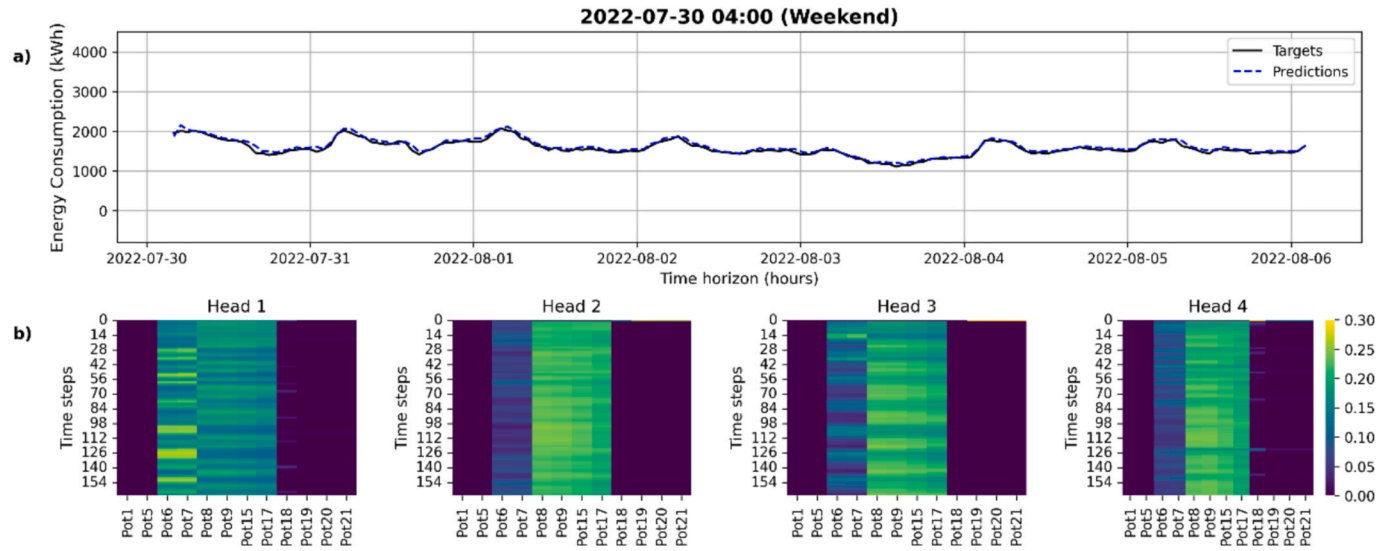

Fig. 9. Representation of the energy demand predictions and attention mechanisms for the selected model by 30-07-2022 at 04:00 (Weekend): (a) target and predicted values of energy demand in kWh; (b) four attention heads of the cross-attention mechanism (encoder–decoder).

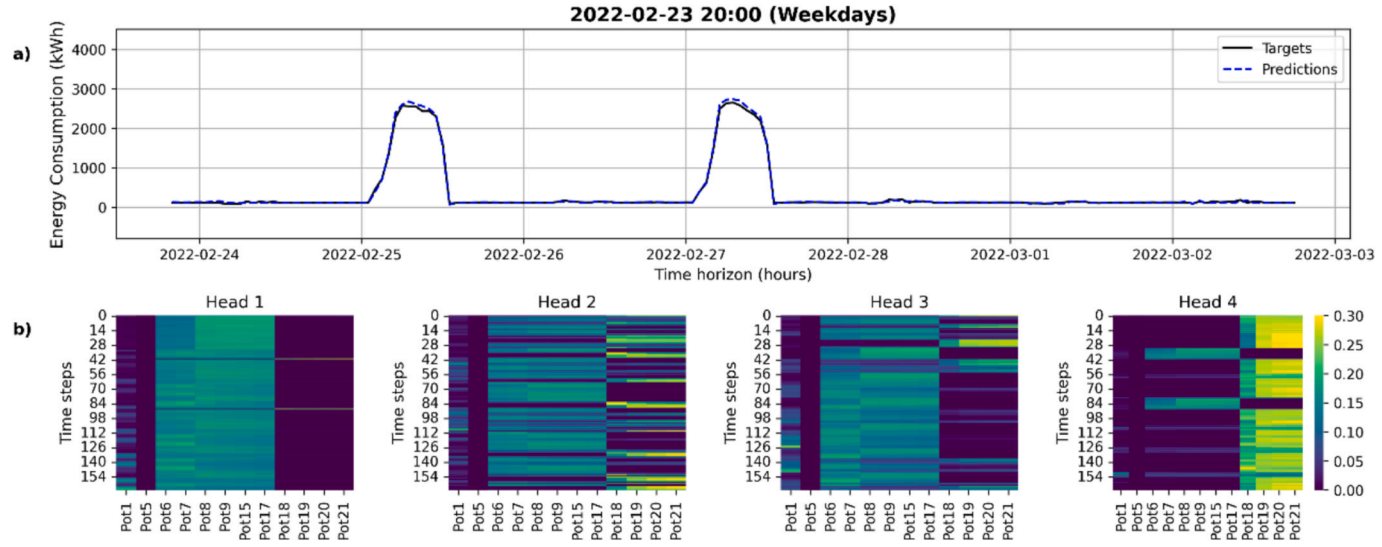

Fig. 10. Representation of the energy demand predictions and attention mechanisms for the selected model by 23-02-2022 at 20:00 (Weekdays): (a) target and predicted values of energy demand in kWh; (b) attention heads of the cross-attention mechanism (encoder–decoder).

relevant. In such contexts, temporal variables have limited influence, as irrigation typically follows a consistent pattern during the months of highest water demand. In contrast, during non-irrigation periods (Figs. 10 and 11), the model shifts its attention patterns. Notably, head 4 alternates its focus between meteorological variables during consumption peaks and recent *ED* history ($PotI_{18}$ to $PotI_{21}$) during off-peak hours. In early-season periods with steady consumption, attention remains more aligned with climatic variables. However, at the end of the season (Fig. 11), where *ED* patterns are more irregular, the model prioritizes recent consumption hours, especially during the night.

Moreover, temporal variables such as $PotI_1$ (DOY) and $PotI_5$ (daylight hours) are interpreted contextually by the TNN model. At the beginning of the season, the day of the year directly influences *ED* due to limited agronomic activity and favourable weather conditions, which reduce the need for pumping. At the end of the season, particularly during daylight hours, sunlight duration becomes more relevant in heads 2, 3, and 4. This is associated with local irrigation practices, where farmers tend to irrigate at night to take advantage of lower electricity prices in tariff period 6, even though they are billed under a demand-based system.

4.3.3. *Autoregressive dependencies in attention-based forecasting*

Finally, several attention heads, such as heads 1, 3, and 4 in Fig. 9, and heads 2 and 3 in Fig. 10, exhibit a strong dependency on *ED* during the three hours preceding the first prediction (t-1, t-2, and t-3). This reflects the model's autoregressive logic, where early predictions rely on recent values to stabilize the sequence. As the prediction progresses, attention shifts toward exogenous variables such as time, day type, and seasonality, allowing the model to adapt its output to the agronomic and climatic context of each period.

The interpretability offered by these attention maps also provides practical insight for irrigation managers. By identifying which variables drive the forecast at different times of the season, decision-makers can understand how climatic and operational conditions affect energy demand and photovoltaic generation potential. For instance, during periods of high solar radiation, when both irrigation and photovoltaic

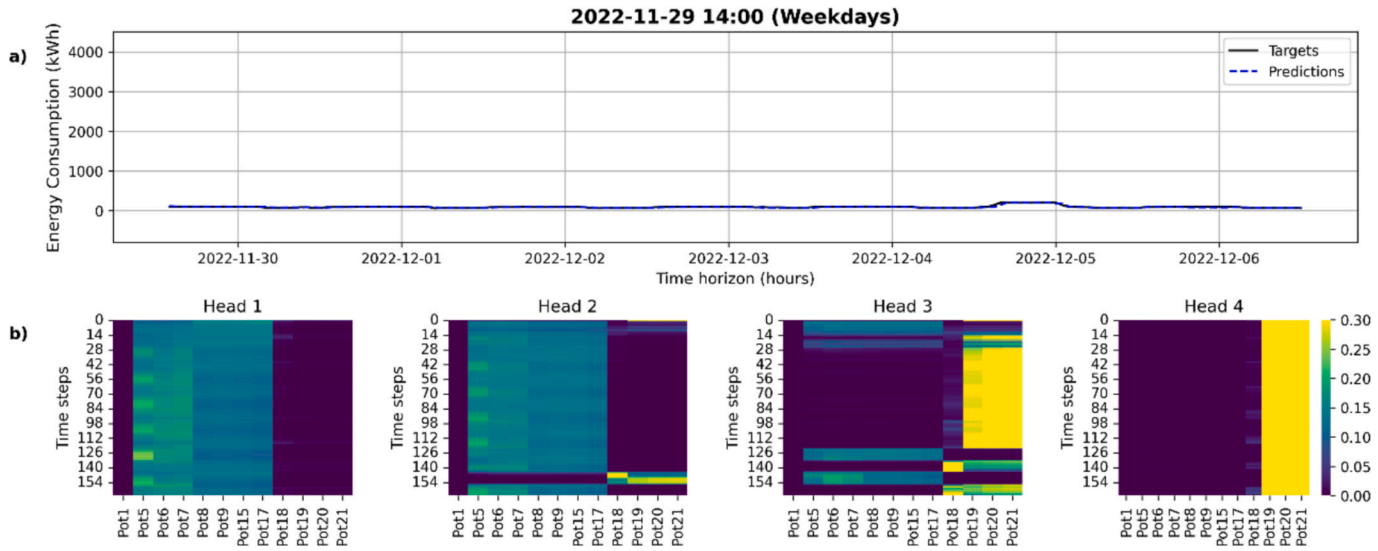

Fig. 11. Representation of the energy demand predictions and attention mechanisms for the selected model by 29-11-2022 at 14:00 (Weekdays): (a) target and predicted values of energy demand in kWh; (b) four attention heads of the cross-attention mechanism (encoder–decoder).

output peak, managers can anticipate possible surpluses or deficits and adjust the energy injection plan accordingly. Thus, the attention-based analysis bridges the gap between data-driven prediction and actionable management, supporting evidence-based decisions that minimize market deviation penalties and promote efficient integration of renewable energy in irrigation systems.

4.3.4. Limitations of the study

While the proposed Transformer-based irrigation demand forecasting model demonstrated strong performance, several considerations should be taken into account. First, the quality and availability of the dataset impose inherent constraints, as missing or noisy data can affect prediction accuracy. Second, computational resources limited the size and complexity of the model that could be tested, including the number of layers, attention heads, and batch size, which may restrict performance in larger-scale implementations. Third, the model is site-specific: its parameters and input transformations are optimized for the studied irrigation district, requiring re-calibration or retraining for other districts. Finally, like all data-driven forecasting models, the Transformer model can only learn from the data it has been trained on; in scenarios significantly different from the training conditions, such as prolonged droughts or sudden changes in irrigation practices, performance may degrade, necessitating model adjustment or retraining. These considerations underscore the importance of local context, data quality, and the temporal scope of training data when deploying such models.

5. Conclusions

This research presents a robust mid-term forecasting model for irrigation *ED*, developed and validated using extensive real-world data from the VIGID over four irrigation seasons (2020–2023). Through rigorous FL and statistical analyses, 12 key input variables were identified, reflecting both temporal and meteorological factors critical to irrigation patterns. The developed TNN, optimized with early stopping and dropout regularization to mitigate overfitting, demonstrated exceptional predictive performance, with an average R2 of 99.62% and low error metrics (RMSE = 0.038, MAE = 32.36 kWh). This confirms the model's capacity to accurately capture complex, nonlinear consumption behaviours and seasonal variability inherent to irrigation systems. The multi-head attention mechanism further enhanced interpretability by dynamically weighting input variables according to context, highlighting the relevance of climatic conditions during peak irrigation periods and temporal factors outside the irrigation season. While prediction accuracy slightly decreased at extreme *ED* values due to limited data density, the model remains highly reliable for operational decision-making and energy management. These results underscore the potential of advanced attention-based architectures in precision agriculture, enabling data-driven optimization of water and energy resources.

Importantly, the selected input variables include universally applicable factors, such as climatic conditions (e.g., air temperature and solar radiation), temporal indicators, and energy price signals, which are widely available across most regions. These drivers are not exclusive to irrigation systems but are common determinants of energy demand for a broad range of water- and energy-intensive end users. This supports the potential generalizability of the proposed framework beyond irrigation districts to other water and energy demand contexts. While local recalibration may be required to account for site-specific operational practices, infrastructure characteristics, and market conditions, the underlying modelling approach and variable selection strategy are transferable and scalable.

In summary, the proposed hybrid model demonstrates significant improvements in hourly energy demand forecasting for ID, providing valuable support for operational decision-making and resource management. Ongoing and future research will explore its application to other water-related energy end-users, such as aquaculture facilities, energy communities, and port infrastructures, as well as the systematic evaluation of model reliability across diverse operational and geographical settings. Furthermore, future work may include the development of models capable of predicting renewable energy production and integrating predictive models into optimization algorithms to efficiently allocate and manage available resources.

CRediT authorship contribution statement

Mariana Akemi Ikegawa Bernabé: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Rafael González Perea:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Conceptualization. **Juan Antonio Rodríguez Díaz:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition. **Jorge García Morillo:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition.

Funding

This research was supported by HY4RES – Hybrid Solutions for Renewable Energy Systems: achieving net-zero Atlantic area energy consumers & communities), grant number EAPA_0001/2022, from Interreg Atlantic Area Programme and co-funded by the European Union.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Valle Inferior del Guadalquivir Irrigation District for providing part of the dataset used in this study.

Data availability

The authors do not have permission to share data.

References

Abed, M., Imteaz, M.A., Ahmed, A.N., Huang, Y.F., 2022. A novel application of transformer neural network (TNN) for estimating pan evaporation rate. Appl Water Sci 13, 31. https://doi.org/10.1007/s13201-022-01834-w.

Allam, T., Mcewen, J.D., 2023. Paying Attention to Astronomical Transients: Introducing the Time-series Transformer for Photometric Classification.

Antonesi, G., Cioara, T., Anghel, I., Papias, I., Michalakopoulos, V., Sarmas, E., 2025. Hybrid transformer model with liquid neural networks and learnable encodings for buildings' energy forecasting. Energy AI 20, 100489. https://doi.org/10.1016/J.EGYAI.2025.100489.

Behzadipour, F., Ghasemi Nezhad Raeini, M., Abdanan Mehdizadeh, S., Taki, M., Khalil Moghadam, B., Zare Bavani, M.R., Lloret, J., 2023. A smart IoT-based irrigation system design using AI and prediction model. Neural Comput. Appl. 35, 24843–24857. 10.1007/s00521-023-08987-y.

Bernabé, M.A.I., Chacón, M.C., Díaz, J.A.R., Montesinos, P., Morillo, J.G., 2025. Comparing strategies for optimal pumps as turbines selection in pressurised irrigation networks using particle swarm optimisation: application in canal del Zújar Irrigation District, Spain. Technologies (Basel). 13. https://doi.org/10.3390/technologies13060233.

Cayuela, J.A.F., García, A.M., García, I.F., Díaz, J.A.R., 2024. Life cycle assessment of large-scale solar photovoltaic irrigation. Sci. Total Environ. 954, 176813. https://doi.org/10.1016/J.SCITOTENV.2024.176813.

Comunidad de Regantes del Valle Inferior del Guadalquivir [WWW Document], n.d. URL https://valleinferior.es/ (accessed 3.17.25).

Corder, G.W., Foreman, D.I., 2009. Nonparametric statistics for non-statisticians: A step-by-step approach. Nonparametric Statistics for Non-Statisticians. 12–37. https://doi.org/10.1002/9781118165881.ch2.

Corominas, J., 2010. Agua y energía en el riego, en la época de la sostenibilidad. Ingeniería Del Agua 17, 219–233. https://doi.org/10.4995/ia.2010.2977.

Cureton, E.E., 1956. Rank-biserial correlation. Psychometrika 21, 287–290. https://doi.org/10.1007/BF02289138.

Red de Información Agroclimática de Andalucía (RIA) | Instituto de Investigación y Formación Agraria y Pesquera (IFAPA) [WWW Document], n.d. URL https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/ (accessed 3.6.25).

Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W., 2017. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 74, 902–924. https://doi.org/10.1016/J.RSER.2017.02.085.

Dutilleul, P., Stockwell, J.D., Frigon, D., Legendre, P., 2000. The Mantel test versus Pearson's correlation analysis: Assessment of the differences for biological and environmental studies. J. Agric. Biol. Environ. Stat. 5, 131–150. https://doi.org/10.2307/1400528.

González Perea, R., Camacho Poyato, E., Montesinos, P., Rodríguez Díaz, J.A., 2019a. Optimisation of water demand forecasting by artificial intelligence with short data sets. Biosyst. Eng. 177, 59–66. https://doi.org/10.1016/j.biosystemseng.2018.03.011.

González Perea, R., Camacho Poyato, E., Montesinos, P., Rodríguez Díaz, J.A., 2019b. Prediction of irrigation event occurrence at farm level using optimal decision trees. Comput. Electron. Agric. 157, 173–180. https://doi.org/10.1016/J.COMPAG.2018.12.043.

González Perea, R., Camacho Poyato, E., Rodríguez Díaz, J.A., 2024. Attention is all water need: Multistep time series irrigation water demand forecasting in irrigation disctrics. Comput. Electron. Agric. 218. https://doi.org/10.1016/j.compag.2024.108723.

González Perea, R., Fernández García, I., Camacho Poyato, E., Rodríguez Díaz, J.A., 2023. New memory-based hybrid model for middle-term water demand forecasting in irrigated areas. Agric Water Manag 284, 108367. https://doi.org/10.1016/J.AGWAT.2023.108367.

Hahnloser, R.H.R., Sarpeshkar, R., Mahowald, M.A., Douglas, R.J., Seung, H.S., 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405, 947–951. https://doi.org/10.1038/35016072.

Horstkötter, E., Zhu, Q., 2025. Computational Experimental Test on PID Controlled Fixed Wing Aircraft Systems. ICCK Transactions on Sensing, Communication, and Control 2, 95–105. 10.62762/TSCC.2025.731885.

Hussan, U., Wang, H., Peng, J., Jiang, H., Rasheed, H., 2026. Transformer-based renewable energy forecasting: A comprehensive review. Renew. Sustain. Energy Rev. 226, 116356. https://doi.org/10.1016/J.RSER.2025.116356.

Iqbal, M., Yousaf, J., Khan, A., Muhammad, T., 2025. IoT-Enabled Food Freshness Detection Using Multi-Sensor Data Fusion and Mobile Sensing Interface. ICCK Transactions on Sensing, Communication, and Control 2, 122–131. 10.62762/TSCC.2025.401245.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://doi.org/10.1016/J.COMPAG.2018.02.016.

Kavya, M., Mathew, A., Shekar, P.R., P, s.,, 2023. Short term water demand forecast modelling using artificial intelligence for smart water management. Sustain. Cities Soc. 95. https://doi.org/10.1016/j.scs.2023.104610.

Lin, T., Wang, Y., Liu, X., Qiu, X., 2022. A survey of transformers. AI Open 3, 111–132. https://doi.org/10.1016/J.AIOPEN.2022.10.001.

Lin, Y., Cunningham, G.A., Coggeshall, S.V., 1996. Input variable identification — fuzzy curves and fuzzy surfaces. Fuzzy Set. Syst. 82, 65–71. https://doi.org/10.1016/0165-0114(95)00223-5.

Liu, S., Ding, F., 2025. Iterative Estimation Algorithm for Bilinear Stochastic Systems by Using the Newton Search. ICCK Transactions on Intelligent Systematics 2, 76–84. 10.62762/TIS.2024.155941.

Nair, V., Hinton, G., 2010. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair, Proceedings of ICML.

Nascimento, E.G.S., de Melo, T.A.C., Moreira, D.M., 2023. A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy. Energy 278, 127678. https://doi.org/10.1016/J.ENERGY.2023.127678.

OMI-Polo Espanol S.A. [WWW Document], 2025. URL https://www.omie.es/en/market-results/ (accessed 1.14.25).

Rigneault, H., Kumar, N.G., Cossart, R., Septier, D., Brévalle-Waslilewki, G., Kudlinski, A., Kaszas, A., 2020. Attention mechanism, transformers, BERT, and GPT: Tutorial and survey. focus on. Microscopy 255, 10.34894/VQ1DJA.

Rodríguez-Díaz, J.A., Pérez-Urrestarazu, L., Camacho-Poyato, E., Montesinos, P., 2011. The paradox of irrigation scheme modernization: More efficient water use linked to higher energy demand. Span. J. Agric. Res. 9, 1000–1008. https://doi.org/10.5424/sjar/20110904-492-10.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0.

Saeed, F., Rehman, A., Shah, H.A., Diyan, M., Chen, J., Kang, J.M., 2025. SmartFormer: Graph-based transformer model for energy load forecasting. Sustainable Energy Technol. Assess. 73, 104133. https://doi.org/10.1016/J.SETA.2024.104133.

Sheikh Khozani, Z., Barzegari Banadkooki, F., Ehteram, M., Najah Ahmed, A., El-Shafie, A., 2022. Combining autoregressive integrated moving average with long short-term memory neural network and optimisation algorithms for predicting ground water level. J. Clean. Prod. 348, 131224. https://doi.org/10.1016/J.JCLEPRO.2022.131224.

Tarjuelo, J.M., Rodriguez-Diaz, J.A., Abadía, R., Camacho, E., Rocamora, C., Moreno, M.A., 2015. Efficient water and energy use in irrigation modernization: Lessons from Spanish case studies. Agric Water Manag 162, 67–77. https://doi.org/10.1016/j.agwat.2015.08.009.

Tian, F., Fan, X., Wang, R., Qin, H., Fan, Y., 2022. A power forecasting method for ultra-short-term photovoltaic power generation using transformer model. Math. Probl. Eng. 2022. https://doi.org/10.1155/2022/9421400.

van de Loo, M., Camacho Poyato, E., van Halsema, G., Rodríguez Díaz, J.A., 2024. Defining the optimization strategy for solar energy use in large water distribution networks: A case study from the Valle Inferior irrigation system. Spain. Renew. Energy 228, 120610. https://doi.org/10.1016/J.RENENE.2024.120610.

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in. Adv. Neural Inf. Proces. Syst. 10.48550/arXiv, 1706.03762.

Wang, J.Q., Du, Y., Wang, J., 2020. LSTM based long-term energy consumption prediction with periodicity. Energy 197. https://doi.org/10.1016/j.energy.2020.117197.

Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Appears Neural Computation 1, 270–280. https://doi.org/10.1162/neco.1989.1.2.270.

Wolfert, S., Isakhanyan, G., 2022. Sustainable agriculture by the internet of things – a practitioner's approach to monitor sustainability progress. Comput. Electron. Agric. 200, 107226. https://doi.org/10.1016/J.COMPAG.2022.107226.

Woo, G., Liu, C., Sahoo, D., Kumar, A., Hoi, S., 2022. ETSformer: Exponential Smoothing Transformers for Time-series Forecasting.

Wu, H., Xu, J., Wang, J., Long, M., 2022. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting.

Xiao, C., Ye, J., Esteves, R.M., Rong, C., 2016. Using Spearman's correlation coefficients for exploratory data analysis on big dataset. Concurr. Comput. 28, 3866–3878. https://doi.org/10.1002/cpe.3745.

Yan, K., Li, W., Ji, Z., Qi, M., Du, Y., 2019. A hybrid LSTM neural network for energy consumption forecasting of individual households. IEEE Access 7, 157633–157642. https://doi.org/10.1109/ACCESS.2019.2949065.

Zeng, M., Deng, X., 2025. Topic Mining and Sentiment Analysis for Consumer Reviews of Automotive Spare Parts on E-commerce Platforms. ICCK Transactions on Intelligent Systematics 2, 137–148. 10.62762/TIS.2025.106283.

Zhang, M., Wang, R., 2025. Energy consumption forecasting in non-stationary machining processes based on a physics-guided transformer. Comput. Ind. 171, 104321. https://doi.org/10.1016/J.COMPIND.2025.104321.

Zhao, J., Chu, F., Xie, L., Che, Y., Wu, Y., Burke, A.F., 2026. A survey of transformer networks for time series forecasting. Comput. Sci. Rev. 60, 100883. https://doi.org/10.1016/J.COSREV.2025.100883.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W., 2020. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: 35th AAAI Conference on Artificial Intelligence. https://doi.org/10.1609/aaai.v35i12.17325.

Glossary

ADAM: Adaptive Moment Estimation
AI: Artificial Intelligence
ANNs: Artificial Neural Networks
ARIMA: Autoregressive Integrated Moving Average
CHG: Guadalquivir River Basin Authority
CNN: Convolutional Neural Networks
d_{ff}*:* Dimension of the FNN layer
DL: Deep Leaning
DOY: Days of year
ED: Energy Demand (kWh)
ETo: Evapotranspiration
FL: Fuzzy Logic
FNN: Feed-forward Network
GA: Genetic Algorithms
GRU: Gated Recurrent Unit
ID: Irrigation District
IV_x*:* Input variable x of the model
K: Key vector
LSTM: Long Short-Term Memory
MAE: Mean Absolute Error
ML: Machine Learning
MSE: Mean Squared Error
NLP: Natural Language Processing
PBC: Point-Biserial Correlation
PCC: Pearson Correlation Coefficient
PE: Positional Encoding
$Potl_X$*:* Variable input potential X
Q: Query vector
R^2*:* Coefficient of Determination
ReLU: Rectified Linear Unit
RMSE: Root Mean Squared Error
SCM: Spearman Correlation Matrix
$t \pm h$*:* Current ED and h hour(s) before or after
TNN: Transformer Neural Networks
V: Value vector