

Reliability Index Analysis for Evaluating Input Variables' Impact on Pump Speed and Efficiency in Drinking Water Wells

Jamshid Piri¹✉

¹ Associate Professor, Department of Water Engineering, University of Zabol, Zabol, Iran.
Corresponding Author, email: J.piri@uoz.ac.ir

Received: 01 April 2025

Accepted: 05 May 2025

Available online: 10 May 2025

Abstract: In this study, a convolutional neural network (CNN) is used to evaluate pump performance and reliability in drinking water wells. A study of 17 wells in Bandar Torkaman, Iran, reveals significant seasonal variation in pump efficiency and operation. The CNN architecture uses convolutional layers with 32 and 64 filters to transform time-series pump data into two-dimensional representations for feature extraction. March showed moderate accuracy (MABE = 26.82, RMSE = 65.70, $R^2 = 0.78$), while July showed the highest error rates (MABE = 171.60, RMSE = 386.9, $R^2 = 0.54$). Model performance peaked in January (MABE = 14.60, RMSE = 39.4, $R^2 = 0.95$) and February (MABE = 48.13, RMSE = 83.20, $R^2 = 0.81$). In 10 of 12 months the Weibull distribution best characterized water levels, while the lognormal distribution dominated flow rate patterns. Model accuracy showed a strong seasonal dependence, with optimal performance during the winter months ($R^2 > 0.80$) and reduced accuracy during summer and early autumn ($R^2 < 0.50$). The analysis revealed that CNN performance correlated closely with seasonal variation in pump operation patterns, which suggests the need for seasonally adjusted prediction methods. These findings provide insight for optimizing pump operations in drinking water wells, particularly in regions with pronounced seasonal variation, and highlight the importance of considering temporal factors in predictive modeling.

Keywords: Pump optimization; Monte Carlo simulation; Reliability index; Artificial neural network; Genetic algorithm.

1. Introduction

Managing water resources has become increasingly important in the twenty-first century, particularly in arid and semi-arid regions where water scarcity poses significant challenges for sustainable development and public health. Despite their central role in water supply infrastructure, pumping systems in drinking water wells face growing operational challenges related to efficiency, reliability, and energy consumption. Rahman et al. (2024) report that optimizing these systems could reduce operational costs by up to 30 % while improving supply reliability. Water pumping systems consume approximately 10–20 % of global electrical energy, with significant potential for efficiency improvement (Zakariazadeh et al. 2024; Piri et al. 2023), and this share can reach 25–30 % of water-sector energy use in developing countries (Kumar et al. 2022). Optimizing these systems therefore presents compelling opportunities for energy conservation and operational cost reduction while ensuring a reliable water supply for communities (Gómez-Ramos et al. 2024). Analytical approaches are required to deal with the complex interaction between the various operational parameters of water pumping systems. In regions with pronounced seasonal variation in water demand and availability, traditional methods of pump system analysis fail to capture the dynamic nature of these interactions. Kardani et al. (2022) have documented significant advances in pump operation optimization using probabilistic methods and artificial intelligence techniques. The present research builds on these advances by combining the two.

1.1. Current state of research

Energy efficiency and reliability analysis have been the primary foci of recent pump system optimization studies, but rarely both at the same time. Coppola et al. (2003) and Esen et al. (2008) showed that artificial neural networks could predict pump performance under different operating conditions with an accuracy of 95 % (Esen et al. 2008; Coppola Jr et al. 2003; Roy et al. 2021). Singh et al. (2024) noted that their work did not address the reliability issues associated with pump operation in any depth. Piri et al. (2021) took a similar approach when developing reliability models for wastewater pumping systems, but did not fully integrate energy efficiency considerations into their calculations. Several recent studies have demonstrated the effectiveness of Monte Carlo simulation for assessing the reliability of pump systems. Asani et al. (2023) advanced these techniques by integrating machine learning approaches with the methodologies developed by De Benedictis et al. (2023). Using Monte Carlo methods combined with artificial intelligence, Jin et al. (2022) achieved 25 % energy savings by optimizing pump systems.

1.2. Theoretical framework

The reliability analysis of pumping systems operates within the broader framework of systems engineering and probability theory (Dutta et al. 2022; Hussain et al. 2024). Li et al. (2022) identified the reliability index as an important metric for evaluating system performance. Nayak et al. (2023) showed that recent advances in machine learning have created the potential for real-time pump optimization. That work, together with the findings of Pyzer-Knapp et al. (2022), shows that combining artificial neural networks with Monte Carlo

simulation provides a robust framework for analyzing complex system behavior.

1.3. Research context

This research fills a critical gap in pump system optimization by developing an integrated approach that combines reliability analysis with energy efficiency considerations. We evaluate the impact of various input parameters on pump speed and efficiency in drinking water wells, with a focus on seasonal variation. By employing Monte Carlo simulation and artificial neural networks, we introduce a comprehensive framework for real-time pump performance prediction, a seasonal reliability index that accounts for temporal fluctuation in system performance, and a practical decision-support approach for optimizing pump operation under varying environmental conditions. This methodology extends beyond traditional single-parameter optimization and offers both theoretical advances in water resource management and practical solutions for improving pump system efficiency and reliability in arid regions.

2. Materials and Methods

2.1. Study area

This study focuses on 17 water wells in the Bandar Torkaman region, a coastal city in Golestan Province, Iran. These wells are the primary source of drinking water for the urban and rural populations of Bandar Torkaman and Gomishan. Bandar Torkaman is situated on the southern coast of the Caspian Sea and has a diverse ecosystem shaped by its proximity to both the sea and inland wetlands. The geographical coordinates of the study area are given in Table 1.

Table 1. The geographical coordinates of the study area.

Coordinate	Value
Latitude	36° 54' 35" N
Longitude	54° 06' 49" E
Elevation	-20 m

The region is a flat, low-lying coastal plain and is home to several small islands, including Ashuradeh (Figure 1).

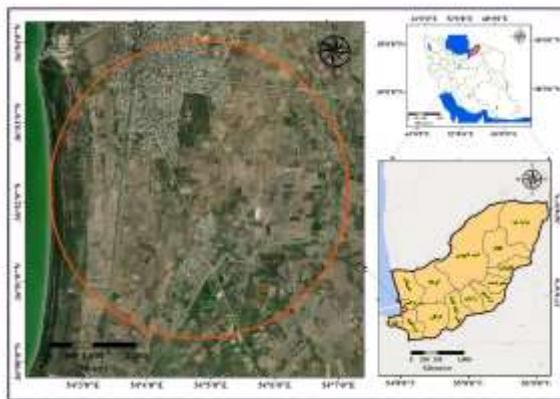


Figure 1. Map of the study area.

A methodological framework for analyzing pump performance in drinking water wells is illustrated in Figure 2. Five parameters are measured in the 17 water

wells: well depth, pump installation depth, well discharge, pump speed, and water level. Four probability distributions are first fitted to the data (normal, lognormal, gamma, and Weibull), and model selection criteria are applied to determine the best-fitting distribution. Two parallel analytical branches then follow: CNN modeling and reliability analysis. The CNN model consists of an input layer, convolutional layers, a max-pooling layer, fully connected layers, and an output layer, while the reliability branch calculates the reliability index through Monte Carlo simulation and yields a seasonal performance evaluation. Both branches converge to provide a comprehensive evaluation of pump performance and reliability. This integrated approach combines deep learning with probabilistic reliability assessment to provide a robust framework for analyzing and optimizing drinking water well systems.

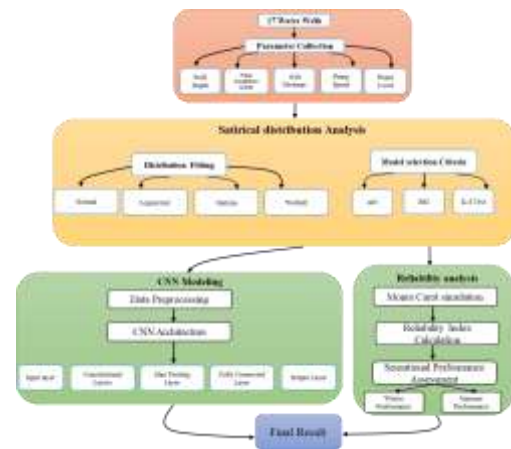


Figure 2. Methodology framework for CNN-based reliability analysis of water well systems.

2.2. Climatic characteristics

Bandar Torkaman has a hot and humid climate owing to its proximity to the Caspian Sea. Summers are characterized by high temperatures and humidity, while winters are relatively mild and humid. The city has an average annual temperature of approximately 20 °C and an annual precipitation of 600 to 700 mm, concentrated mainly in autumn and winter. The soils of the region are predominantly sandy and clayey and support coastal and wetland vegetation.

2.3. Hydrogeological characteristics

The Bandar Torkaman region is underlain by coastal aquifers fed by groundwater. Proximity to the Caspian Sea and the wetlands results in varying water quality within these aquifers, whose thickness ranges from 10 to 30 m. The bedrock of the region consists primarily of permeable alluvial and sandy deposits, which facilitate groundwater recharge. The water table is relatively high because of the influence of the Caspian Sea and the wetlands, and it fluctuates seasonally in response to rainfall. Although the water table is generally higher than in southern areas such as Kordkoy, the water tends to be more saline.

2.4. Main research data

Data were collected from 17 wells over a four-year period (2020–2023). The dataset includes the following variables.

- **Well depth.** Accurate measurements of well depth were used to calculate static head.
- **Pump installation depth.** This parameter was recorded to assess its impact on pumping capacity and pump protection.
- **Well discharge.** Discharge rates varied seasonally and with water demand.
- **Pump speed.** Pump speed was monitored to evaluate pump performance under varying conditions of discharge and water level.
- **Water level.** Groundwater levels were measured to assess temporal fluctuation.

3. Methodology

3.1. Monte Carlo simulation for reliability analysis

Monte Carlo simulation evaluates pump reliability through repeated random sampling. The method generates probability distributions for key performance metrics using (Piri et al. 2021)

$$P(x) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right) e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

The reliability index (β) is calculated as

$$\beta = \frac{\mu - R}{\sigma} \quad (2)$$

where μ is the mean of the simulated values, R is the reference value (typically the design or target RPM), and σ is the standard deviation of the simulated values. The simulation proceeds in four steps.

Step 1. Generate N random samples ($N = 10,000$) for the input variables,

$$X = \{x_1, x_2, \dots, x_n\} \sim f(X) \quad (3)$$

where $f(x)$ is the fitted probability distribution.

Step 2. For each sample, calculate the performance function,

$$g(x) = RPM_{actual} - RPM_{required} \quad (4)$$

Step 3. Estimate the failure probability,

$$P_f = P(g(x) < 0) = \frac{1}{N} \sum I[g(x_i) < 0] \quad (5)$$

where $I[\cdot]$ is the indicator function.

Step 4. Calculate the monthly reliability index,

$$\beta_{month} = -\Phi^{-1}(P_f) \quad (6)$$

where Φ^{-1} is the inverse standard normal cumulative distribution function. This approach quantifies uncertainty in pump performance and enables risk-based decision making for operational optimization.

3.2. Convolutional neural network modeling for pump performance prediction

The CNN architecture transforms time-series pump data into a two-dimensional representation for feature extraction. Given an input sequence $X = \{x_1, x_2, \dots, x_n\}$

containing water level, flow rate, and month, the convolution operation is (Kiakojoouri 2024)

$$y_i = (x \times w)_i + b = \sum_k x_{i+k} \times w_k + b \quad (7)$$

where w represents the filter weights and b is the bias term. The model structure consists of five layers.

Layer 1. Input transformation,

$$x \in \mathbb{R}^d \rightarrow x' \in \mathbb{R}^{d \times m} \quad (8)$$

where d is the number of features (three) and m is the sequence length.

Layer 2. Convolution and activation, with 32 filters and a 3×3 kernel,

$$h_1 = ReLU(w_1 \times x + b_1) \quad (9)$$

followed by 64 filters and a 3×3 kernel,

$$h_2 = ReLU(w_2 \times h_1 + b_2) \quad (10)$$

Layer 3. Max pooling with 2×2 windows,

$$p(x) = \max\{x_{ij} : i, j \in [1, 2]\} \quad (11)$$

Layer 4. Fully connected layers, with 128 and 64 neurons,

$$FC_1 = ReLU(w_3 h_2 + b_3) \quad (12)$$

$$FC_m = ReLU(w_4 FC_1 + b_4) \quad (13)$$

Layer 5. Output,

$$RPM = w_5 FC_m + b_5 \quad (14)$$

A dropout probability of $p = 0.25$ is applied between layers to prevent overfitting,

$$P(h_i') = \begin{cases} \frac{h_i}{1-p} & \text{with probability } 1-p \\ 0 & \text{with probability } p \end{cases} \quad (15)$$

This architecture exploits the ability of a CNN to detect temporal patterns in pump operation through hierarchical feature extraction, which improves RPM prediction accuracy compared with traditional neural networks.

3.3. Training details and statistical analysis

The hidden layers use the sigmoid activation function,

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (16)$$

while the output layer is linear, $f(x) = x$. Performance was measured with the mean absolute bias error, the root mean square error, and the coefficient of determination,

$$MABE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (18)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (19)$$

3.4. Statistical distribution analysis

Table 2 lists the probability density functions fitted to the data and the criteria used to select among them.

Table 2. Probability density functions and model selection criteria.

Distribution / criterion	Formula	Parameters
Probability distributions		
Normal	$f(x) = (1/(\sigma\sqrt{2\pi})) \exp[-(x - \mu)^2 / (2\sigma^2)]$	μ = mean, σ = standard deviation
Lognormal	$f(x) = (1/(x\sigma\sqrt{2\pi})) \exp[-(\ln x - \mu)^2 / (2\sigma^2)]$	μ = log-mean, σ = log-standard deviation
Gamma	$f(x) = (1/(\Gamma(k)\theta^{k*})) x^{k*-1} e^{-x/\theta}$	k = shape parameter, θ = scale parameter
Weibull	$f(x) = (k/\lambda)(x/\lambda)^{k*-1} e^{-(x/\lambda)^k}$	k = shape parameter, λ = scale parameter
Model selection criteria		
Akaike information criterion	$AIC = -2 \ln(L) + 2k$	L = likelihood, k = number of parameters
Bayesian information criterion	$BIC = -2 \ln(L) + k \ln(n)$	n = sample size
Kolmogorov–Smirnov test	$D = \sup F(x) - F_n(x) $	$F(x)$ = theoretical CDF, $F_n(x)$ = empirical CDF
Best fit criteria		
Minimum values	Min(AIC, BIC)	Lower values indicate a better fit
Maximum values	Max(K–S p -value)	Higher p -values indicate a better fit

4. Results and Discussion

4.1. Statistical analysis

Four probability distributions (normal, lognormal, gamma, and Weibull) were fitted to the data. The Weibull distribution showed the best fit for water levels in 10 of 12 months, while the lognormal distribution dominated flow rate patterns in seven months. As Table 3 shows, operational data from the 17 wells displayed distinct patterns across the three key parameters of RPM, water level, and flow rate. During the early months (March and April) the gamma distribution provided the best fit, with low AIC and BIC values (210–215) and high K–S p -values approaching unity. From May to June the distribution shifted to lognormal while the goodness of fit remained high (K–S p -value = 0.98) as temperatures increased. AIC and BIC values (227–230) and K–S p -values (0.08–0.71) became more variable in the middle of the year (July to September), which suggests that operational variability had increased.

Water level measurements were remarkably consistent, with the Weibull distribution providing the best fit in 10 of 12 months. The AIC and BIC values of these fits (385–860) were higher than those of the RPM distributions, which suggests more complex behavior. Despite the consistency in distribution type, K–S p -values for the water level fits were generally lower (0.1–0.4). Normal distributions provided better fits in

September and March, possibly reflecting specific hydrological conditions during those months.

Flow rate analysis showed that the lognormal distribution was the dominant pattern and was optimal for seven months of the year. AIC and BIC values for these fits were lower (80–1010) than those for RPM and water level, which suggests simpler underlying patterns. K–S p -values varied between 0.2 and 0.6, indicating moderate fit quality. Gamma distributions held the best fit in March and October and normal distributions in July, reflecting seasonal variation in water consumption patterns. A notable correspondence was observed between the distributions of RPM and flow rate during certain months: both were gamma in March and lognormal in May and June, which suggests a strong relationship between pump performance and water flow. Among the three parameters, RPM showed the greatest variability in optimal distribution type, indicating greater sensitivity to operational and environmental factors. Flow rate showed moderate stability with notable seasonal variation, while water level was the most stable. Based on four years of operational data, these findings provide insight for managing and optimizing drinking water wells. They highlight the complex interactions between pump performance, water levels, and flow rates, and they point to the value of seasonal adjustments to operational strategies.

Table 3. Statistical summary of the 17 wells for the four years 2020–2023.

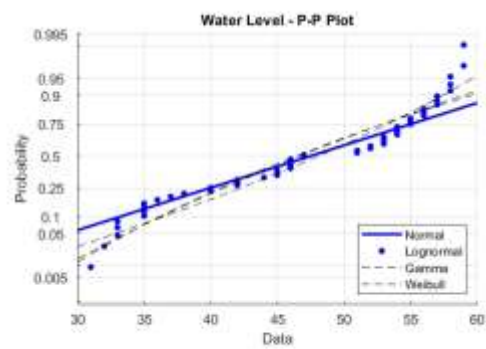
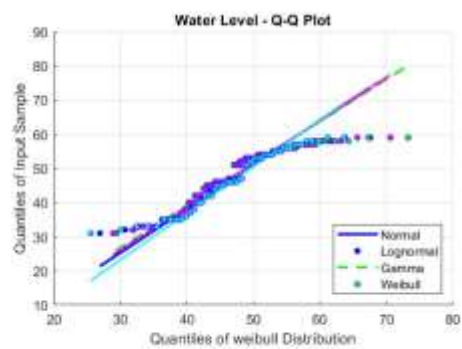
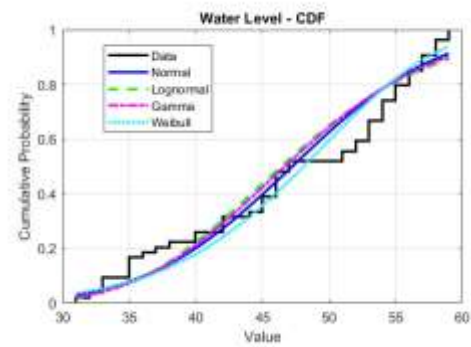
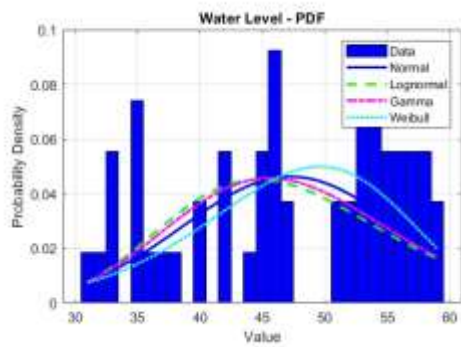
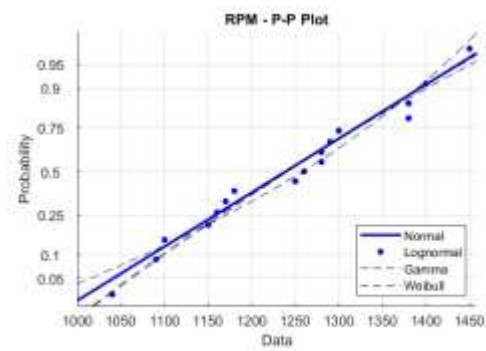
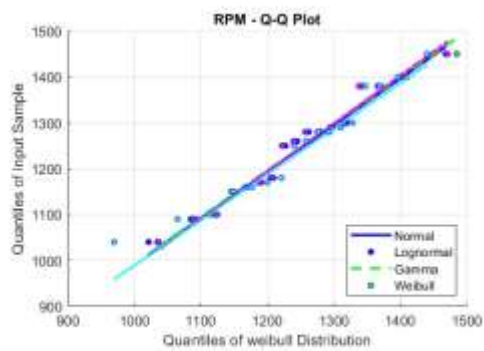
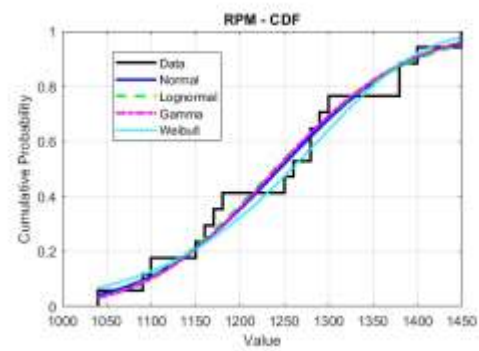
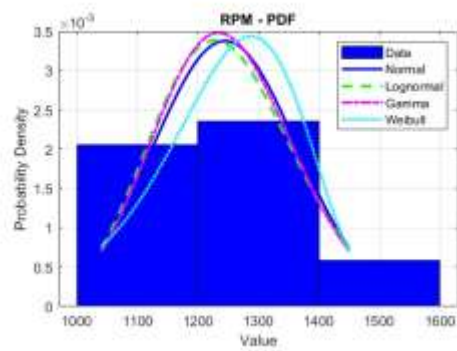
Month	Parameter	Optimal distribution	AIC	BIC	K–S p -value
March	RPM	Gamma	213.48	215.15	0.95
	Water Level	Weibull	385.12	389.10	0.18
	Qp	Gamma	91.78	95.76	0.41
April	RPM	Gamma	207.79	209.46	1.00
	Water Level	Weibull	385.78	389.76	0.34
	Qp	Lognormal	98.12	102.09	0.52
May	RPM	Lognormal	209.69	211.36	0.98
	Water Level	Weibull	401.44	405.45	0.42
	Qp	Lognormal	98.46	102.48	0.44
June	RPM	Lognormal	217.54	219.21	0.98

	Water Level	Weibull	389.72	393.77	0.09
	Qp	Lognormal	82.70	86.82	0.21
July	RPM	Weibull	230.35	232.01	0.34
	Water Level	Weibull	459.74	463.76	0.11
	Qp	Normal	105.40	109.42	0.36
August	RPM	Weibull	227.15	228.82	0.08
	Water Level	Weibull	396.33	400.38	0.39
	Qp	Lognormal	103.56	107.61	0.12
September	RPM	Weibull	228.79	230.45	0.71
	Water Level	Normal	393.75	397.80	0.10
	Qp	Lognormal	88.37	92.43	0.53
October	RPM	Normal	212.46	214.12	0.98
	Water Level	Weibull	394.56	398.61	0.24
	Qp	Gamma	93.36	97.41	0.43
November	RPM	Lognormal	218.83	220.49	0.98
	Water Level	Weibull	402.94	407.06	0.13
	Qp	Lognormal	82.70	86.82	0.21
December	RPM	Gamma	213.48	215.15	0.95
	Water Level	Weibull	415.29	419.41	0.13
	Qp	Lognormal	91.96	96.08	0.58
January	RPM	Normal	213.50	215.16	0.94
	Water Level	Weibull	418.18	422.30	0.13
	Qp	Lognormal	105.08	109.20	0.30
January	RPM	Normal	213.50	215.16	0.94
	Water Level	Normal	422.19	426.31	0.13
	Qp	Lognormal	105.08	109.20	0.30

RPM = revolutions per minute; Qp = pumping flow rate.

As shown in Figure 3, MATLAB was used to perform a comprehensive statistical analysis of the well data for March (2020–2023). Four common statistical distributions (normal, lognormal, gamma, and Weibull) were fitted to the RPM, water level, and flow rate data, and probability density functions, cumulative distribution functions, Q–Q plots, and P–P plots were generated. The results revealed distinct distribution patterns. The RPM data fitted the gamma distribution best (selected three times), which indicates asymmetric pump performance patterns influenced by environmental and operational factors. The water level

data aligned best with the Weibull distribution (selected four times), which suggests the influence of erosional factors and the gradual changes typical of natural phenomena. The flow rate data fitted the lognormal distribution best (selected twice), which indicates positive skewness and the presence of occasional larger values. These findings offer insight for practical applications ranging from predictive modeling and operational optimization to maintenance scheduling and resource management, including forecasting low water periods and improving pump maintenance strategies.



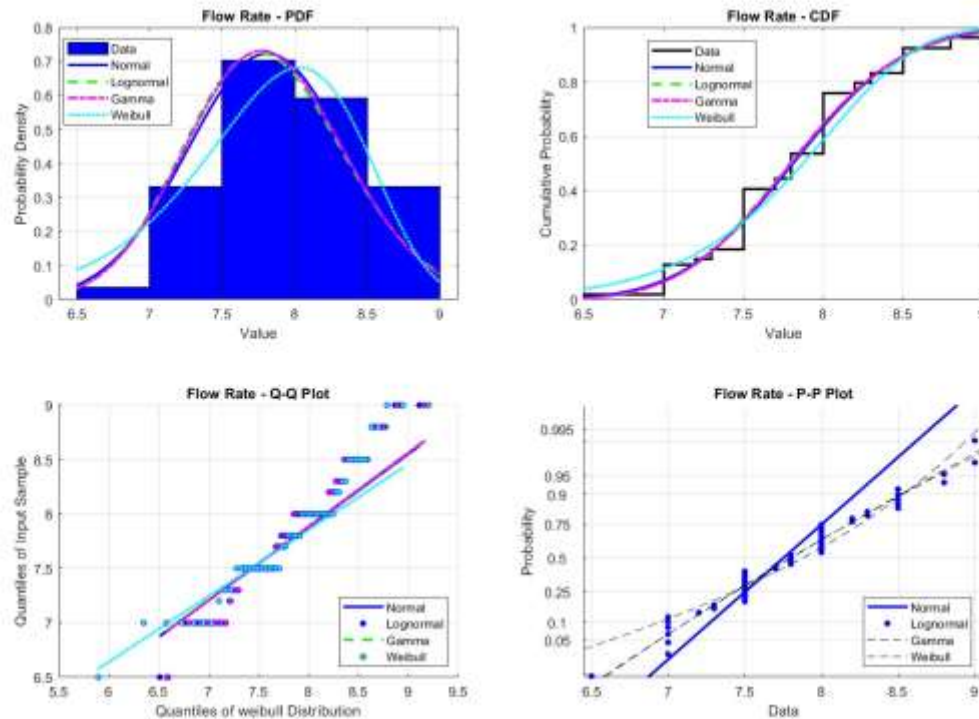
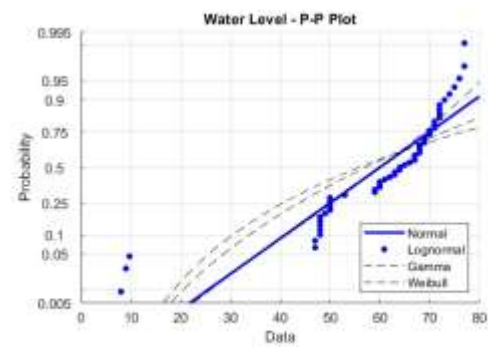
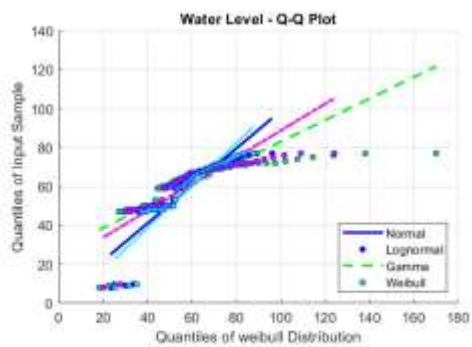
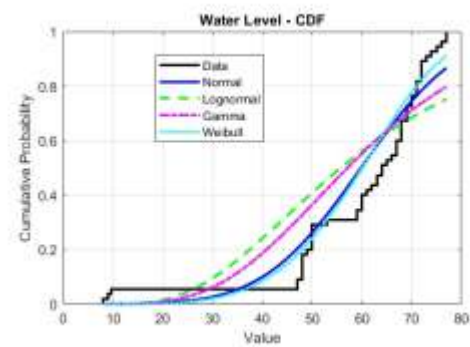
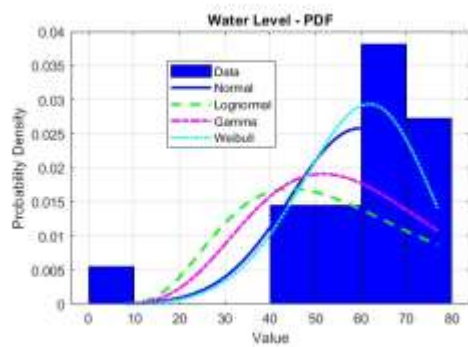
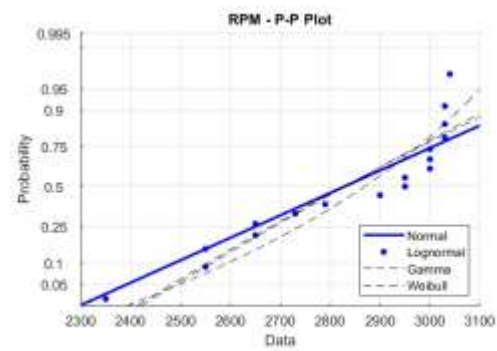
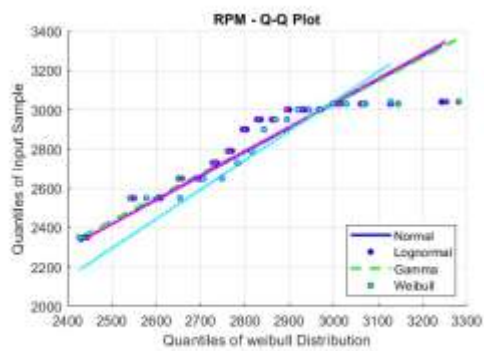
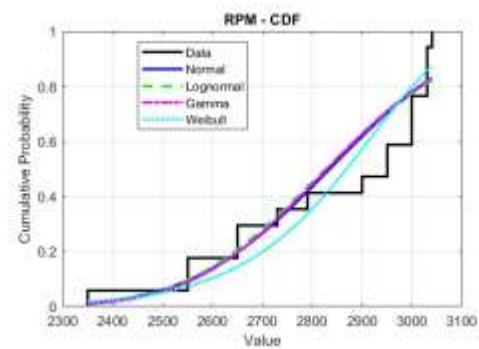
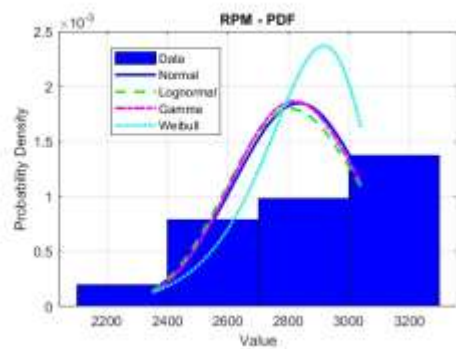


Figure 3. CDF, PDF, Q-Q and P-P plots for March (2020–2023) for the seventeen wells: RPM (top), water level (middle) and flow rate (bottom).

Figure 4 shows the corresponding statistical patterns for July (2020–2023). The lognormal distribution showed the best fit for RPM, being selected four times. The PDF plot for RPM indicates a distribution with positive skewness, which matches the characteristics of a lognormal distribution and may reflect increased pump activity during this hot month in response to peak summer demand. The Q-Q plot for RPM also aligns well with the reference line, which confirms the choice of the lognormal distribution. Water level is best modelled by the Weibull distribution, selected three times. The PDF plot shows a multimodal distribution of water levels, which could indicate the presence of different well groups with varying depths or different hydrological conditions, and the CDF plot shows the gradual increase expected of a Weibull distribution. The lognormal distribution was the best fit for the flow rate

data, selected three times. The flow rate PDF plots are symmetric with a slight right skew, consistent with a lognormal distribution; this slight skew may indicate an overall increase in flow rate during the month, likely because of elevated water consumption in peak summer. The Q-Q and P-P plots for flow rate also fit the lognormal distribution well, although some deviation is observed in the extreme values. These findings suggest that well behavior in July is influenced strongly by increased temperature and water demand, which is evident in the patterns of increased pump activity, significant variation in water level, and elevated flow rates. This information is important for effective water resource management during peak summer, since it allows planning for increased demand and helps to ensure the sustainability of groundwater resources.



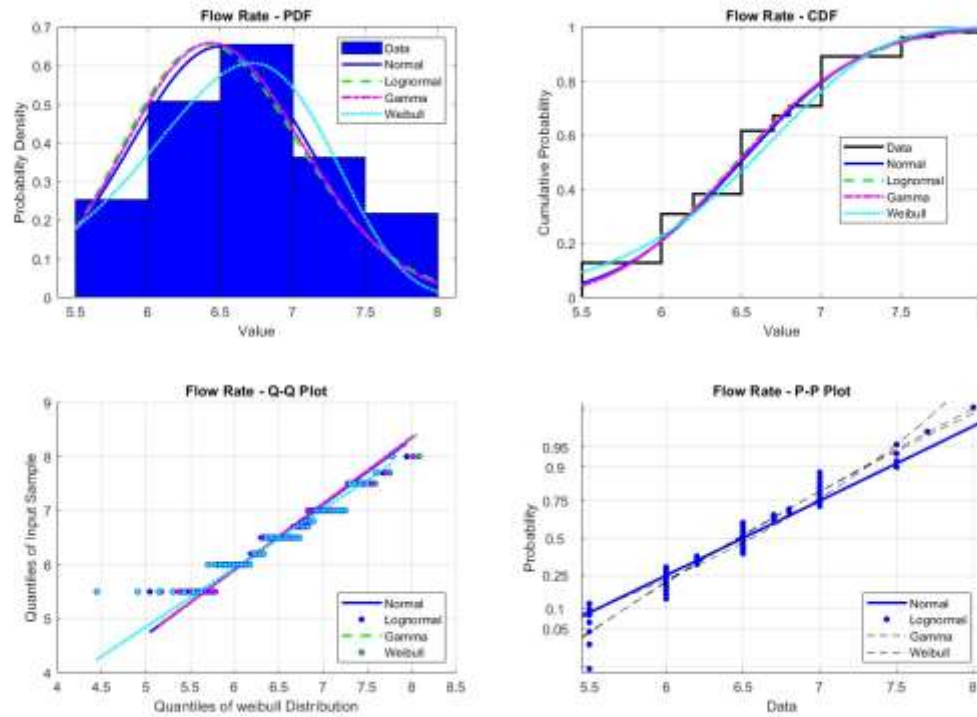
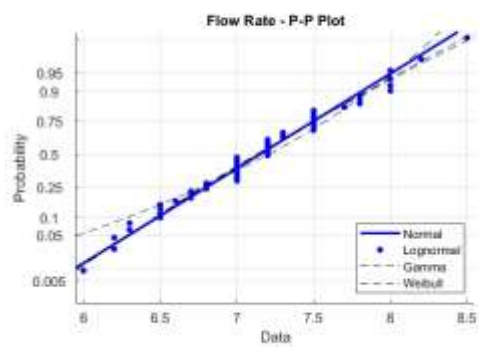
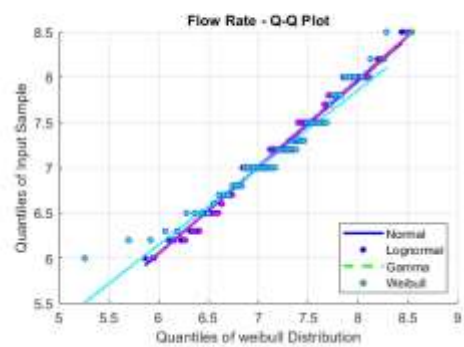
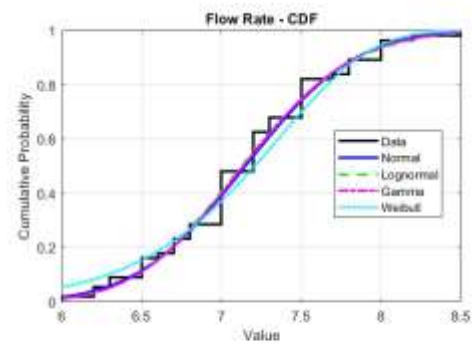
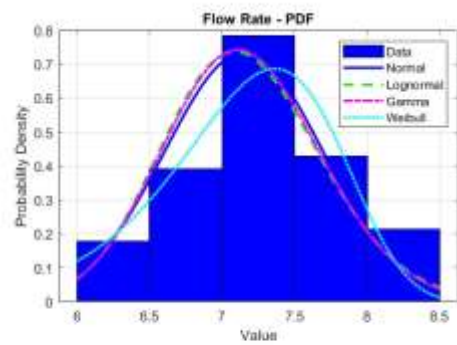
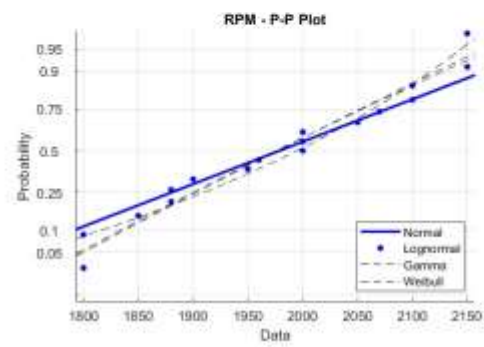
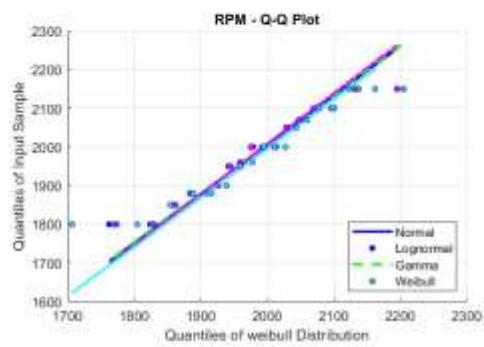
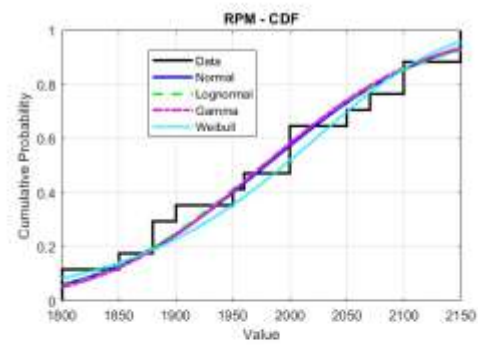
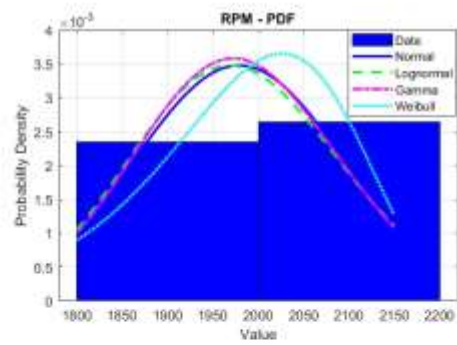


Figure 4. CDF, PDF, Q–Q and P–P plots for July (2020–2023) for the seventeen wells: RPM (top), water level (middle) and flow rate (bottom).

The plots for November (2020–2023) in Figure 5 reveal further significant statistical patterns. For RPM, the Weibull distribution consistently shows the best fit, being selected four times, and the RPM PDF plot shows the positive skewness consistent with Weibull characteristics. Pump performance may vary with the onset of the rainy season, and some high values likely indicate wells still in high demand. The Q–Q plot for RPM aligns well with the reference line, which confirms the choice of the Weibull distribution. The Weibull distribution also provides the best fit for water level, selected three times. The water level PDF plot shows a multimodal distribution, which indicates that autumn rainfall affects different wells differently, and the CDF plot shows a gradual but non-uniform increase in water levels that confirms the complexity of the water level

distribution. For flow rate, the lognormal distribution clearly shows the best fit, selected four times. The flow rate PDF plot shows a slight right skewness similar to lognormal characteristics, which could indicate an overall increase in flow rate due to autumn rainfall while some higher values are maintained. The Q–Q and P–P plots for flow rate align very well with the lognormal distribution. These plots indicate that well behavior in November is influenced by autumn conditions, which results in more complex water level and flow rate patterns. This information is valuable for assessing the impact of autumn rainfall on groundwater resources, for adjusting pumping strategies to changing conditions, and for planning water resource management for the coming cold seasons.



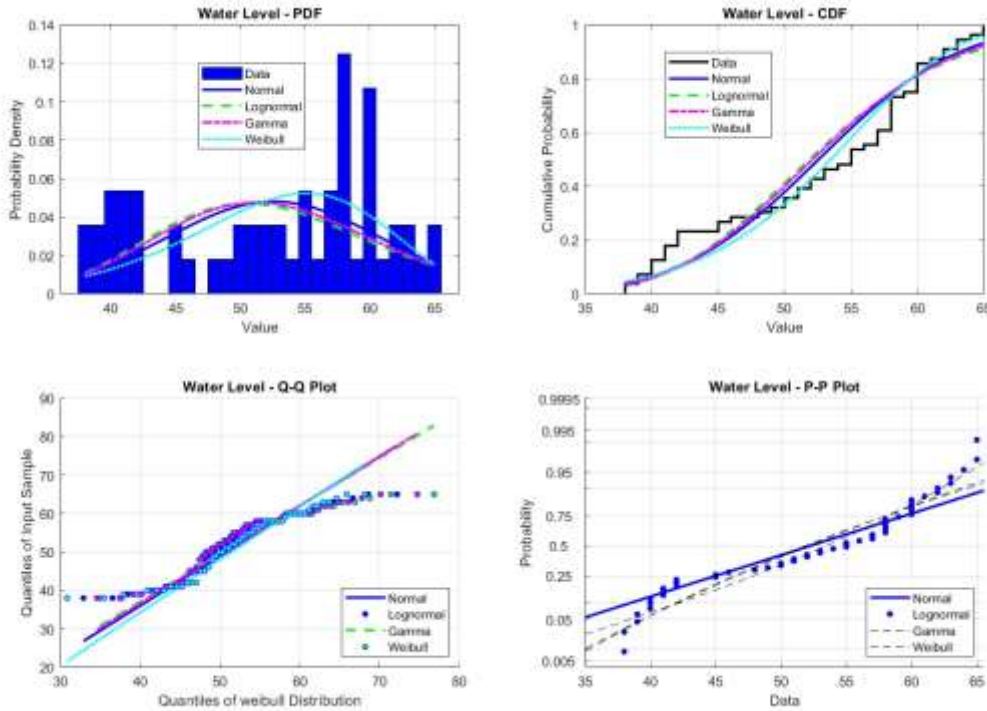


Figure 5. CDF, PDF, Q–Q and P–P plots for November (2020–2023) for the seventeen wells: RPM (top), water level (middle) and flow rate (bottom).

A corresponding statistical analysis of the plots for February (2020–2023) for the seventeen wells reveals significant statistical patterns as well. The gamma distribution shows the best fit for RPM, being selected three times, and the RPM PDF plot shows the positive skewness that matches gamma characteristics. This pattern may indicate greater variation in pump performance during peak winter, caused by differences in freezing conditions and by the need to compensate for reduced flow rates. The Q–Q plot for RPM aligns relatively well with the reference line, although some deviation is observed at high and low values. For water level the Weibull distribution offers the best fit, selected four times; the PDF plot shows a multimodal distribution, which indicates that winter precipitation and snowmelt affect different wells differently, and the CDF plot shows the corresponding gradual but non-uniform increase. Flow rate is best fitted by the lognormal distribution, selected four times, and the positive skewness of the flow rate PDF plot aligns with lognormal characteristics. This could indicate variability in well flow rates, likely because of differing susceptibility to winter conditions and fluctuating demand. The Q–Q and P–P plots for flow rate fit the lognormal distribution very well. These results indicate that well behavior in February is influenced strongly by winter conditions, which produces significant variation in RPM, complex water level patterns, and notable

changes in flow rate. This information is valuable for optimizing water resource management during peak winter, for adjusting pumping strategies to cold and variable conditions, and for planning for the challenges of freezing and of sudden changes in water supply and demand.

4.2. CNN model performance

As shown in Figure 6, the performance of the CNN model exhibited significant seasonal variation. Optimal accuracy was achieved in February, with a substantial decline in performance during autumn (September and October). Model performance fluctuated during spring and summer, with peak accuracy in June and July. Analysis of the RPM range revealed distinct seasonal patterns, with wider ranges in summer and early autumn indicating increased operational variability, and a narrower range in late autumn and winter suggesting more stable well conditions or reduced activity. Seasonal adjustments are therefore crucial to improving overall model performance, since accounting for the specific characteristics of each season can enhance predictive capability significantly. The R^2 metric supports these findings, with the highest value (0.95) in February and the lowest (0.44) in September and October. This highlights the importance of addressing seasonal factors when developing robust machine learning models for predicting well RPM.

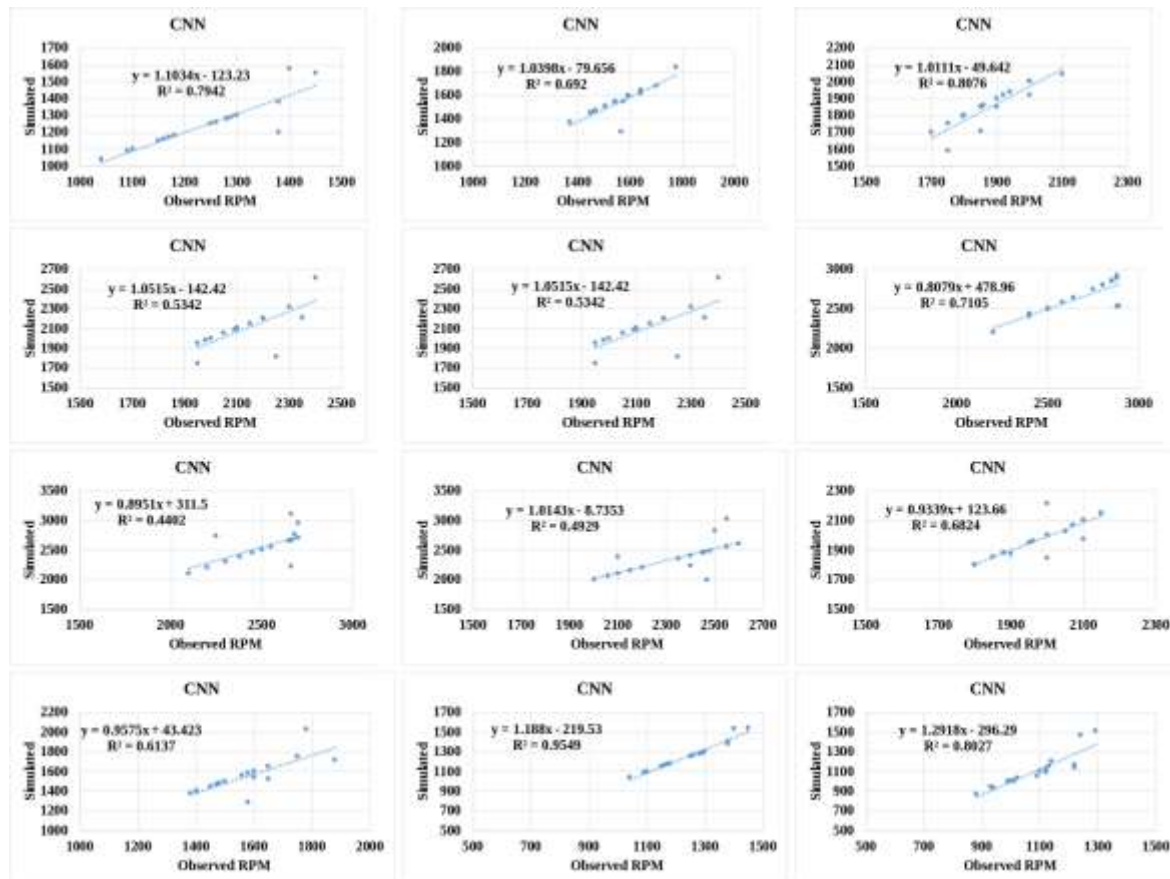


Figure 6. Annual performance of the CNN model in predicting drinking water well RPM. Each panel compares simulated with observed RPM for one month of the year.

The CNN model demonstrated significant seasonal variation in its ability to predict well RPM. The model achieved its highest accuracy during the winter months (January and February), with R^2 values approaching 0.95. Its performance declined substantially during the summer months (June and July), with R^2 values dropping below 0.5 (Table 4). These seasonal patterns suggest that underlying factors such as temperature fluctuation, rainfall, or operational changes influence the RPM values and the ability of the model to capture them. Future research could explore incorporating more detailed meteorological data or developing more complex models to account for these seasonal variations and improve overall performance.

Table 4. Monthly analysis of CNN model performance in predicting well RPM: error and correlation study.

Month	Model	MABE	RMSE	R^2
March	CNN	26.820	65.698	0.78
April	CNN	24.455	69.796	0.69
May	CNN	28.861	58.487	0.80
June	CNN	59.558	132.566	0.53
July	CNN	171.574	386.860	0.54
August	CNN	45.672	121.182	0.44
September	CNN	99.535	200.814	0.49
October	CNN	101.301	197.524	0.68
November	CNN	32.238	71.587	0.68
December	CNN	52.413	106.781	0.61
January	CNN	14.588	39.380	0.95
February	CNN	48.130	83.175	0.80

A violin plot illustrates the monthly error distribution of the CNN model throughout the year (Figure 7). Each violin represents the probability density of the errors at different values, the central horizontal line indicates the median (MABE), and the vertical line indicates the range of error variation (related to RMSE). July has the highest error rate, with a MABE of 171.574 and a substantial spread, followed by September and October with MABE values around 100. The best performance is shown in January, with the lowest MABE (14.588) and the most compact error distribution. A clear seasonal pattern in model accuracy emerges, with the summer months generally showing larger errors and greater variability than the winter months, which suggests that seasonal factors affect model accuracy.

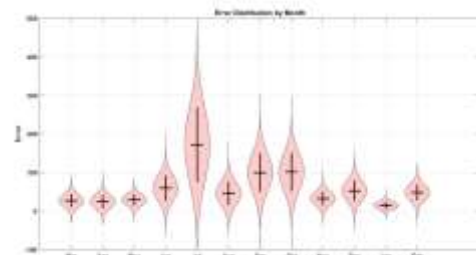


Figure 7. Monthly error distribution pattern of CNN model performance.

Figure 8 compares the RPM distributions predicted by the CNN model with the observed data and the Monte Carlo simulations over a complete year, and reveals significant and diverse patterns. During the first months of the year (March to May) the performance of the CNN is relatively favorable. The model estimates the RPM distribution with acceptable accuracy in March, although it tends to overestimate higher RPM values. Performance improves significantly in April and May, with considerable overlap between the predicted and observed curves. RPM patterns grow more complex as summer approaches. The model continues to perform well in June but faces difficulty in predicting extreme high values, and in July it has the most difficulty capturing the multiple fluctuations in the real data. August and September are the hardest months for the model, which fails to capture the complex distributions and abrupt changes in RPM that characterize them. A gradual improvement can be observed from October

onwards. Although minor differences remain in November and December, the match between the actual data and the CNN predictions improves significantly. January and February are the peak months for model performance: in January the predicted and actual curves align almost perfectly, and February is also strong, with only subtle differences in the details of the distribution. This annual analysis shows that the ability of the CNN to predict the RPM distribution is generally good but is affected by seasonal factors, and that the model performs better in the colder seasons. Seasonal variation should therefore be taken into account when modeling and predicting RPM. More precise parameter tuning, the addition of seasonal variables, and more advanced machine learning techniques are recommended to improve model performance, particularly in months with more complex patterns, so that the model can better capture sudden and severe changes in the RPM distribution.

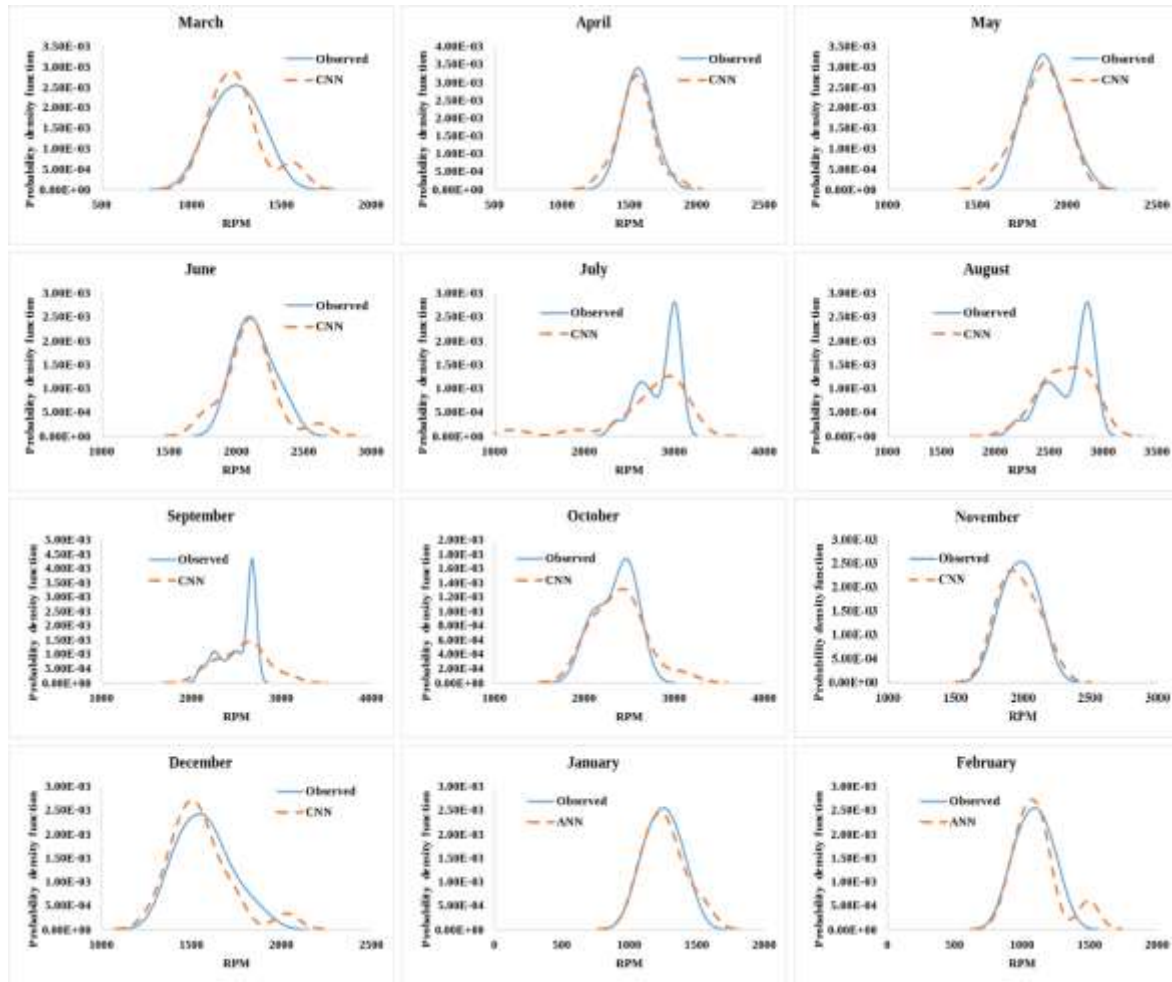


Figure 8. Comparison of observed and CNN-predicted probability distributions of RPM in potable wells.

Monte Carlo analysis of the reliability index (β) shows significant variation in pump performance and energy consumption throughout the year (Figure 9). In winter (January, February, and March) and early spring (April) the index reaches high values (between 1.7185 and 2.1893), which indicates optimal and reliable pump

performance; favorable weather conditions and balanced demand may also make energy consumption more efficient. The index declines sharply during the summer months (June and July) and becomes negative in July and August (-0.0502 and -0.1510), which indicates serious performance challenges for the pumps and

probably increased energy consumption. Higher temperatures and changes in water consumption patterns place the pumping system under greater stress. A gradual improvement in the index is observed from late summer into autumn, which suggests that the system returns progressively to a more favorable state. Specific strategies are therefore required during the warmer months to optimize energy consumption and pump performance, which emphasizes the importance of seasonal management of the pumping system.

5. Conclusions

CNN modeling was applied to predict pump performance in drinking water well systems in Bandar Torkaman, Iran. The model predicts pump performance effectively, but with notable seasonal variation in accuracy: performance peaks in winter ($R^2 = 0.95$ in January) and falls in summer ($R^2 < 0.54$ in July). Several conclusions follow.

The predictive capability of the CNN shows a strong seasonal dependency, which indicates that environmental factors affect pump performance patterns significantly. The winter months show the highest accuracy (MABE = 14.60, RMSE = 39.38), while the summer months present more challenging operational conditions (MABE = 171.60, RMSE = 386.86). Water levels are best characterized by the Weibull distribution across most months, while flow rates are best characterized by the lognormal distribution. These findings can inform future modeling approaches and improve understanding of the underlying behavior of pump systems.

Seasonally adjusted operational strategies for drinking water wells are crucial. Because model performance varies significantly across seasons, pump operation protocols should be adjusted dynamically to account for seasonal changes in water demand, environmental conditions, and system behavior. Future research should focus on improving model performance during the challenging summer period by incorporating additional environmental parameters or developing season-specific models, and hybrid modeling approaches may help to account for seasonal variation in pump performance. These results provide a foundation for improving the efficiency and reliability of drinking water well operations, particularly in regions with pronounced seasonal variation.

References

- Asani PC, Alam Z, Poddar R (2023) Exploring the impact of PVC and PVA microplastics on zebrafish tissue using multi-spectral imaging, Optical Coherence Tomography (OCT) and biospeckle OCT (bOCT). *Chemosphere* 341: 140088.
- Coppola Jr E, Szidarovszky F, Poulton M, Charles E (2003) Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions. *Journal of hydrologic engineering* 8(6): 348-360.
- De Benedictis A, de Palma L, Rossi-Espagnet MC, Marras CE (2023) Connectome-based approaches in pediatric epilepsy surgery: "State-of-the art" and future perspectives. *Epilepsy & Behavior* 149: 109523.
- Dutta N, Kaliannan P, Paramasivam S (2022) A comprehensive review on fault detection and analysis in the pumping system. *International Journal of Ambient Energy* 43(1): 6878-6898.
- Esen H, Inalli M, Sengur A, Esen M (2008) Performance prediction of a ground-coupled heat pump system using artificial neural networks. *Expert Systems with Applications* 35(4): 1940-1948.
- Gómez-Ramos A, Blanco-Gutiérrez I, Ballesteros-Olza M, Esteve P (2024) Are Non-Conventional Water Resources the Solution for the Structural Water Deficit in Mediterranean Agriculture? The Case of the Segura River Basin in Spain. *Water* 16(7): 929.
- Hussain M, Zhang T, Jamil I, Soomro AA, Hussain I (2024) Application of Machine Learning Approaches to Prediction of Corrosion Defects in Energy Pipelines. *Advances in Corrosion Modelling*: 127-166.
- Jin W, Atkinson TA, Doughty C, Neupane G, Spycher N, McLing TL, Dobson PF, Smith R, Podgorney R (2022) Machine-learning-assisted high-temperature reservoir thermal energy storage optimization. *Renewable Energy* 197: 384-397.
- Kardani N, Bardhan A, Samui P, Nazem M, Zhou A, Armaghani DJ (2022) A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil. *Engineering with Computers*: 1-20.
- Kiakojoury A (2024) A Generalised AI-based Model for Rolling Element Bearing Fault Diagnosis in Rotating Machinery. University of Southampton.
- Kumar R, Pandey A, Samykano M, Aljafari B, Ma Z, Bhattacharyya S, Goel V, Ali I, Kothari R, Tyagi V (2022) Phase change materials integrated solar desalination system: An innovative approach for sustainable and clean water production and storage. *Renewable and Sustainable Energy Reviews* 165: 112611.
- Li N, Sun N, Cao C, Hou S, Gong Y (2022) Review on visualization technology in simulation training system for major natural disasters. *Natural hazards* 112(3): 1851-1882.
- Nayak SK, Pradhan B, Mohanty B, Sivaraman J, Ray SS, Wawrzyniak J, Jarzębski M, Pal K (2023) A Review of Methods and Applications for a Heart Rate Variability Analysis. *Algorithms* 16(9): 433.
- Piri J, Abdollahipour M, Keshtegar B (2023) Advanced machine learning model for prediction of drought indices using hybrid SVR-RSM. *Water Resources Management* 37(2): 683-712.
- Piri J, Pirzadeh B, Keshtegar B, Givvehchi M (2021) Reliability analysis of pumping station for sewage network using hybrid neural networks-genetic algorithm and method of moment. *Process Safety and Environmental Protection* 145: 39-51.

- Pyzer-Knapp EO, Pitera JW, Staar PW, Takeda S, Laino T, Sanders DP, Sexton J, Smith JR, Curioni A (2022) Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials* 8(1): 84.
- Rahman MH, Menezes BC, Baldacci R (2024) Exploring the role of blockchain technology, warehouse automation, smart routing, and cloud computing in logistics performance. *Production & Manufacturing Research* 12(1): 2393614.
- Roy DK, Biswas SK, Saha KK, Murad KFI (2021) Groundwater level forecast via a discrete space-state modelling approach as a surrogate to complex groundwater simulation modelling. *Water Resources Management* 35(6): 1653-1672.
- Singh A, Patel S, Bhadani V, Kumar V, Gaurav K (2024) AutoML-GWL: Automated machine learning model for the prediction of groundwater level. *Engineering Applications of Artificial Intelligence* 127: 107405.
- Zakariazadeh A, Ahshan R, Al Abri R, Al-Abri M (2024) Renewable energy integration in sustainable water systems: A review. *Cleaner Engineering and Technology*:100722.