http://communications.science.ankara.edu.tr
ISSN: 1303-6025 E-ISSN: 2651-3749
Commun. Fac. Sci. Univ. Ank. Ser. C Biology
Volume 35, Number 1, Pages 34-51 (2026)
https://doi.org/10.53447/communc.1809445
Research Article

Development of A Machine Learning-Based Predictive Model for *Klebsiella pneumoniae* Biofilm Formation

Nahlah Alalaqi[1], Ergin Murat Altuner[2*]

[1]Department of Biology, Institute of Science, Kastamonu University, Kastamonu, Türkiye
[2]Department of Biology, Faculty of Science, Kastamonu University, Kastamonu, Türkiye

ABSTRACT

Klebsiella pneumoniae is a critical Gram-negative pathogen frequently associated with severe hospital-acquired infections, including pneumonia, urinary tract infections, and bloodstream infections. Its increasing multi-drug resistance, particularly to carbapenems, poses a significant global health challenge. Biofilm formation is a key virulence factor, allowing *K. pneumoniae* to persist on medical devices and evade antibiotic treatment. This study aimed to develop a predictive model to quantify the individual and synergistic effects of key environmental factors, namely temperature, pH, glucose and sodium chloride (NaCl) concentrations, on *K. pneumoniae* biofilm formation. To establish the intricate relationships between these environmental parameters and biofilm development, and to construct a robust predictive model, experimental data were analysed using statistical regression and machine learning approaches to construct a predictive model. Among the tested models, XGBoost demonstrated the best predictive performance (R^2 = 0.6209, Adjusted R^2 = 0.6209, RMSE = 1.2355, MAE = 0.9384). Feature importance analysis revealed temperature and glucose concentration as the most influential factors, significantly impacting biofilm formation. Partial dependence analysis showed optimal biofilm production was observed at neutral pH and glucose concentrations over 1%, with high temperatures. In addition, high NaCl concentrations significantly stimulated biofilm formation. These findings provide valuable insights into the regulation of *K. pneumoniae* biofilm formation, which could have practical uses in diverse industrial, medical, and food safety settings.

Keywords: *Klebsiella pneumoniae*, Biofilm formation, Machine learning, XGBoost, Growth parameters

***Corresponding Author:** Ergin Murat Altuner
awad1996@yahoo.com (N. Alalaqi) 0000-0002-0551-7629 https://ror.org/015scty35
ergin.murat.altuner@gmail.com (E.M. Altuner) 0000-0001-5351-8071 https://ror.org/015scty35

Research Article; Received: October 23, 2025; **Accepted:** March 03, 2026
Cite this article as: Alalaqi, N., Altuner, E.M., Development of a machine learning-based predictive model for *Klebsiella pneumoniae* biofilm formation. *Commun.Fac.Sci.Univ.Ank.Ser.C Biology*, 35 (1) (2026), 34-51. https://doi.org/10.53447/communc.1809445

1 Introduction

Biofilms are intricate microbial environments with microorganisms embedded in an organic polymer matrix. The diverse structural and compositional characteristics of environmental niches, ranging from natural aquatic systems to medical devices, often make their uniform classification challenging. To reflect this extensive variability and broadly

encompass such microbial aggregations, any microbial species immobilised within a matrix is generally termed a 'biofilm' [1]. Biofilms produced by microorganisms in aquatic settings can be advantageous or detrimental. Microorganisms can produce biofilms both in natural and artificial environments [2]. These biofilms are believed to be responsible for 65% of infections in hospitals [3-7]. Biofilms are structured communities embedded in an extracellular polymeric substance (EPS) matrix, which enhances mechanical stability and antimicrobial tolerance [8-11].

Among the most significant biofilm-associated pathogens is *Klebsiella pneumoniae*, a Gram-negative bacterium recognised as a leading cause of severe nosocomial infections, including pneumonia, urinary tract infections, bloodstream infections, and surgical site infections [12,13]. The ability of *K. pneumoniae* to form biofilms is a critical virulence factor, enabling its persistence on medical devices such as catheters and ventilators, and contributing directly to its alarming rates of multi-drug resistance, particularly carbapenem resistance (CRKP) [14-16]. Therefore, understanding the factors influencing *K. pneumoniae* biofilm formation is vital for developing effective control and treatment strategies.

Environmental conditions play a crucial role in regulating bacterial biofilm development [17]. Specifically, temperature, pH, glucose concentration, and sodium chloride (NaCl) concentration are well-known physicochemical parameters that can significantly modulate microbial growth and biofilm architecture [18,19]. While the general impact of these factors on biofilm formation is recognised, their individual and synergistic effects on *K. pneumoniae* biofilm production, and the development of a quantitative predictive model, provide a quantitative framework for evaluating the individual and combined effects of environmental factors on *K. pneumoniae* biofilm formation, supporting data-driven strategies for its control. Additionally, it provides a valuable tool for understanding the complex interactions among these factors, which can inform targeted strategies to combat *K. pneumoniae*-related issues.

Therefore, this study aimed to develop a robust predictive model to assess and quantify the individual and combined effects of temperature, pH, glucose concentration, and sodium chloride concentration on *Klebsiella pneumoniae* biofilm formation. This model is expected not only to quantify the environmental regulation of *K. pneumoniae* biofilm formation but also to bridge experimental microbiology with predictive analytics, introducing a novel data-driven framework for understanding complex microbial behaviours.

2 Method

2.1 Microorganism Used

The *Klebsiella pneumoniae* ATCC BAA-1706 strain was acquired from ThermoScientific and maintained as a freeze-dried stock. For experimental use, the strain was activated by culturing on Luria-Bertani (LB) agar (Merck, Germany) at 37°C for 24 hours and subsequently subcultured twice on fresh LB agar under the same conditions to ensure viability and purity.

2.2 Biofilm Formation Under Varying Environmental Conditions

Biofilm development was carried out using a sterile, U-bottomed 96-well plate. Tryptic Soy Broth (TSB) (Merck, Germany) was used as the culture medium. Biofilm formation was investigated across three pH levels (5, 7, and 9), three sodium chloride concentrations (0%, 3.75%, and 7.5%), and three glucose concentrations (0%, 1%, and 2%). All chemicals, including glucose and sodium chloride, were purchased from Merck, Germany.

The inoculum standardised to 0.5 McFarland ($\approx 1.5 \times 10^8$ CFU/mL) was prepared. For each well, 200 µL of the TSB medium was inoculated with 10 µL of the standardised bacterial suspension. The plates were then incubated for 48 hours at three different temperatures (15°C, 30°C, and 45°C) without agitation, and the medium was not changed throughout the incubation period. Six replicates were used in the research for each condition [20].

2.3 Biofilm Quantification using Crystal Violet Assay

The biofilm formed after incubation was quantified using the crystal violet (CV) assay. Following the 48-hour incubation, the contents of each well were discarded, and the plates were gently washed three times with 200 µL of sterile 0.9% (w/v) NaCl solution per well to remove non-adherent cells. After the final wash, the plates were inverted and air-dried at room temperature for 15 minutes. Subsequently, 200 µL of a 0.1% (w/v) crystal violet solution was added to each well, and the plates were incubated at room temperature for 30 minutes. The excess CV dye was then removed, and the wells were washed three more times with sterile NaCl solution and left to air dry again. To solubilise the bound crystal violet, 200 µL of an ethanol:acetone (70:30) solution was added to each well and allowed to stand for 30 minutes at room temperature. The resulting coloured solution from each well was then transferred to a new, sterile flat-bottom 96-well plate to avoid interference from any residual material in the original plate. The absorbance was measured at 595 nm using a plate reader (BioTek, USA), and values were background-corrected against blank wells [20].

2.4 Optimisation Experimental Design

A fractional factorial design (FFD) (3^{k-1}) was initially employed to screen the influential parameters and investigate their main effects. All experimental runs in the FFD were conducted in six replicates. Recognising the need for a more comprehensive understanding and to model quadratic effects and interactions, a Central Composite Design (CCD) was subsequently implemented. The CCD points were integrated with the FFD runs, allowing for the development of a robust model. Additionally, eleven central point observations were included within the CCD to estimate experimental error and assess curvature.

2.5 Statistical Analysis

All experiments were conducted in six independent replicates (n=6). Statistical analyses were performed using R Studio 2025.09.1 [21].

Initial data screening included an Interquartile Range (IQR) test for outlier detection. Identified outliers were removed in a single iterative step to minimise distortion of data distribution while preserving integrity, ensuring no excessive data trimming occurred. Data normality was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated using Levene's test. As the raw data consistently deviated from a normal distribution, a series of transformations, including logarithmic, square root, inverse, exponential, Box-Cox, Yeo-Johnson, and Z-score, were applied. The Box-Cox transformation was ultimately selected for subsequent analyses due to its superior efficacy in reducing data skewness and kurtosis, as evidenced by achieving the lowest Coefficient of Variation, thereby bringing the data closest to a normal distribution.

Despite the Box-Cox transformation, full adherence to normality or homoscedasticity could not be achieved across all variables. Consequently, non-parametric statistical tests were employed where appropriate. Specifically, the Kruskal-Wallis test ($\alpha = 0.05$) was utilised to identify significant differences among experimental conditions, followed by Dunn's post-hoc test with Bonferroni correction for pairwise comparisons.

To further investigate the main and interaction effects of the independent parameters on the Box-Cox transformed absorbance, a Permutation-based ANOVA was also conducted with 9999 permutations. This non-parametric method was chosen due to the persistent non-normality of the data even after transformation.

The bivariate relationships among variables were explored through Spearman's rank correlation analysis, a non-parametric method selected due to the persistent deviation from full normality in the transformed data. A significance level of $\alpha = 0.05$ was maintained for all inferential statistical tests.

A diverse array of regression models was employed to rigorously investigate the relationship between the independent parameters (temperature, pH, glucose, and NaCl concentrations) and the Box-Cox transformed absorbance (dependent variable). This comprehensive approach aimed to assess both the individual and interactive effects of these factors on *K. pneumoniae* biofilm formation, while also exploring a wide spectrum of model

complexities.

The modelling suite included traditional statistical regression methods such as Linear, Quadratic, Polynomial, Ridge, Lasso, Elastic Net, and Stepwise regression. In parallel, advanced machine learning algorithms, which are particularly adept at capturing non-linear relationships and interactions, were also applied. These encompassed Decision Tree, Random Forest, Gradient Boosting, XGBoost, Light GBM, and CatBoost regression models.

Collectively, these models were utilised to provide a thorough assessment of variable relationships and to construct a robust predictive framework for biofilm absorbance, enabling the identification of the most suitable model for this dataset.

3 Results and Discussion

The individual and combined effects of temperature, pH, glucose, and sodium chloride (NaCl) concentrations on *K. pneumoniae* biofilm formation were rigorously assessed. Non-parametric Kruskal-Wallis tests revealed statistically significant variations in biofilm production across all tested parameters ($p < 0.05$ for each factor). This indicated that at least one level within each environmental factor significantly influenced biofilm development, confirming the initial hypothesis (H_1). Further pairwise comparisons were performed using Dunn's post-hoc test with Bonferroni correction to identify specific differences between groups, which are indicated by different letters in Figure 1, Figure 2, Figure 3 and Figure 4.

3.1 Effect of Temperature

Many mushroom species exhibit antioxidant potential due to the presence of bioactive compounds such as phenolics, flavonoids, and ascorbic acid, which contribute to redox balance and cellular protection [13,14]. In this context, the present study evaluated the antioxidant parameters of *R. chloroides*, and the results are summarized in Table 1.

3.2 Effect of pH

Biofilm formation by *K. pneumoniae* was also significantly influenced by pH values (Figure 2a, 2b, and 2c). pH 7.0 consistently supported higher biofilm biomass across various temperature and nutrient conditions, suggesting it is near the optimal pH for *K. pneumoniae* biofilm development. Both acidic (pH 5.0) and alkaline (pH 9.0) conditions generally led to reduced biofilm formation compared to neutral pH.

3.3 Effect of NaCl Concentration

The concentration of NaCl had a statistically significant impact on *K. pneumoniae* biofilm development (Figure 3a, 3b, and 3c). There are statistically significant differences in biofilm formations in different NaCl concentrations, highlighting the osmoregulatory challenges for *K. pneumoniae* at elevated salinity.

3.4 Effect of Glucose Concentration

K. pneumoniae biofilm development was significantly impacted by glucose concentration (Figure 4a, 4b, and 4c). Biofilm production generally increased with increasing glucose concentrations, with 1% and 2% glucose supporting significantly higher biomass compared to 0% glucose, indicating that glucose serves as a crucial carbon source for biofilm matrix synthesis and bacterial growth. The highest biofilm levels were often recorded at 2% glucose under optimal pH and temperature.

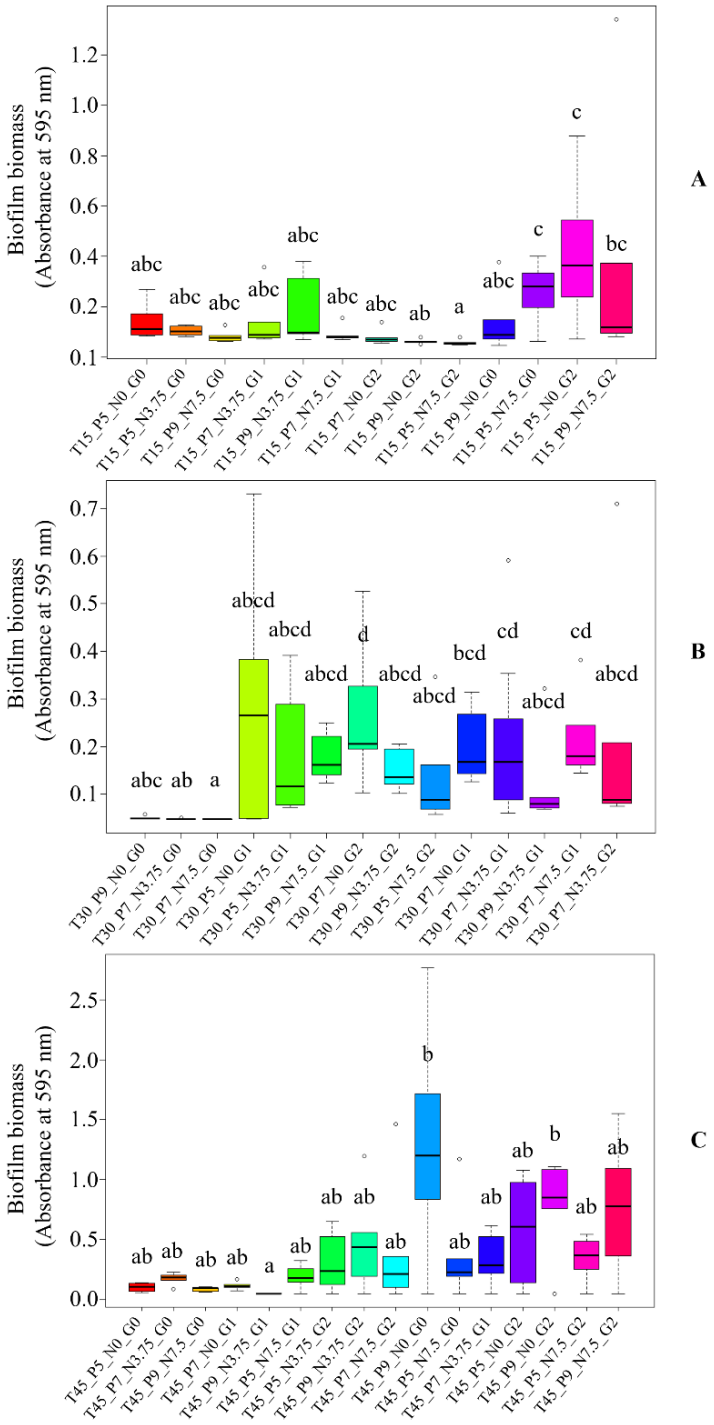

Figure 1: Distribution of absorbances at different temperatures
a) 15°C, b) 30°C, c) 45°C (x-axis = T: Temperature, P: pH, N: NaCl (%), and G: Glucose (%) values)

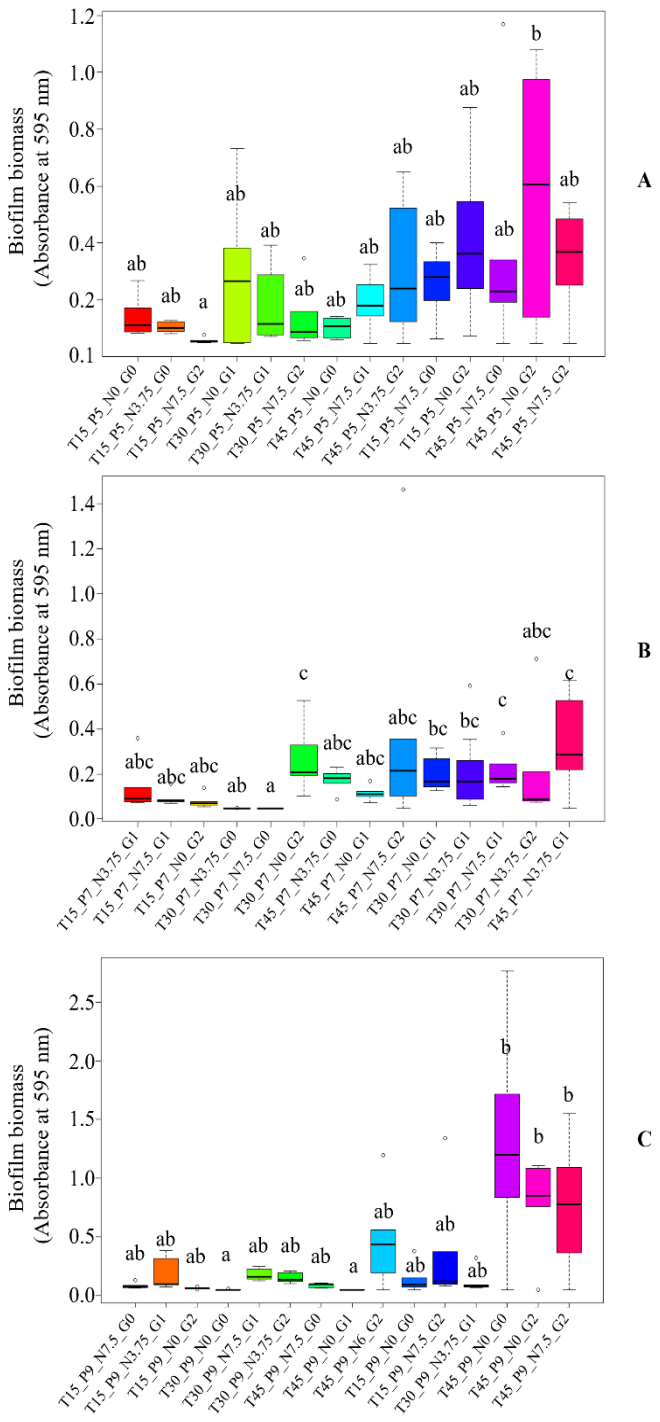

Figure 2: Distribution of absorbances at different pH values
a) pH = 5.0, b) pH = 7.0, c) pH = 9.0 (x-axis = T: Temperature, P: pH, N: NaCl (%), and G: Glucose (%) values)

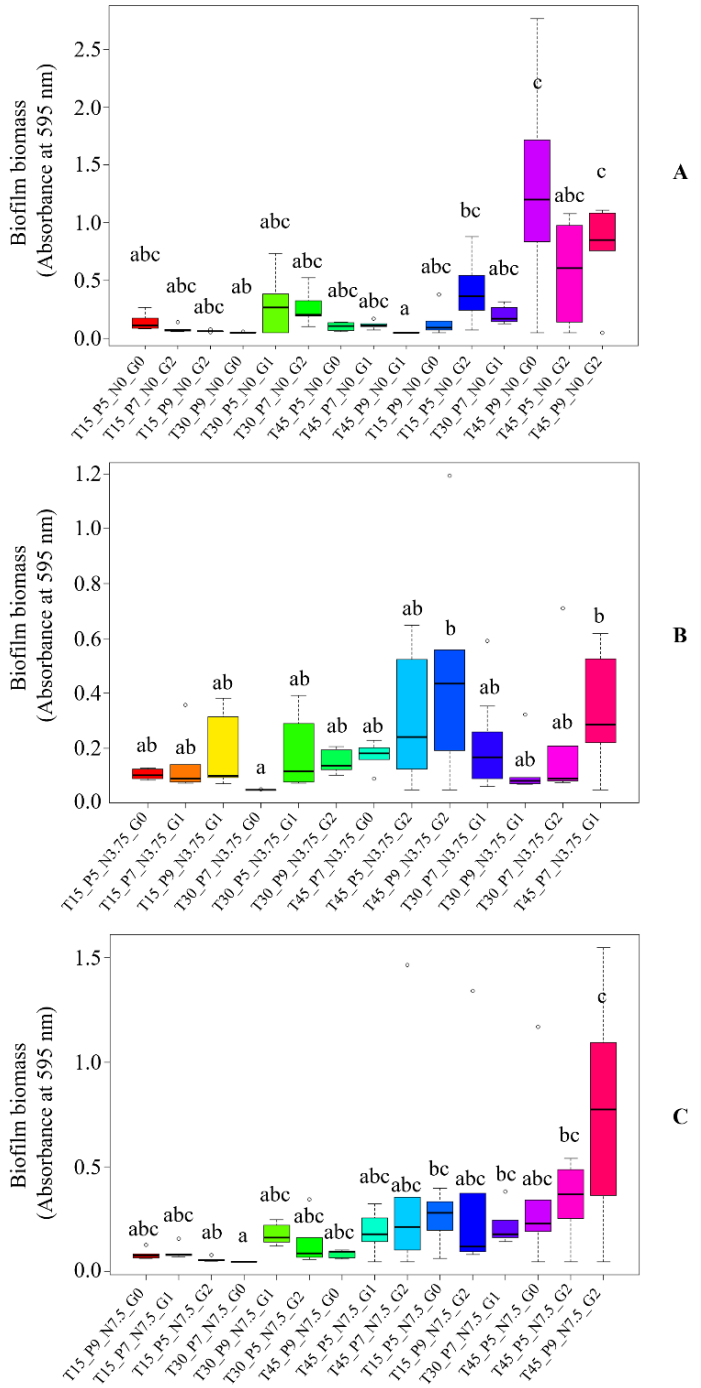

Figure 3: Distribution of absorbances at different NaCl concentrations
a) 0%, b) 3.75%, c) 7.5% (x-axis = T: Temperature, P: pH, N: NaCl (%), and G: Glucose (%) values)

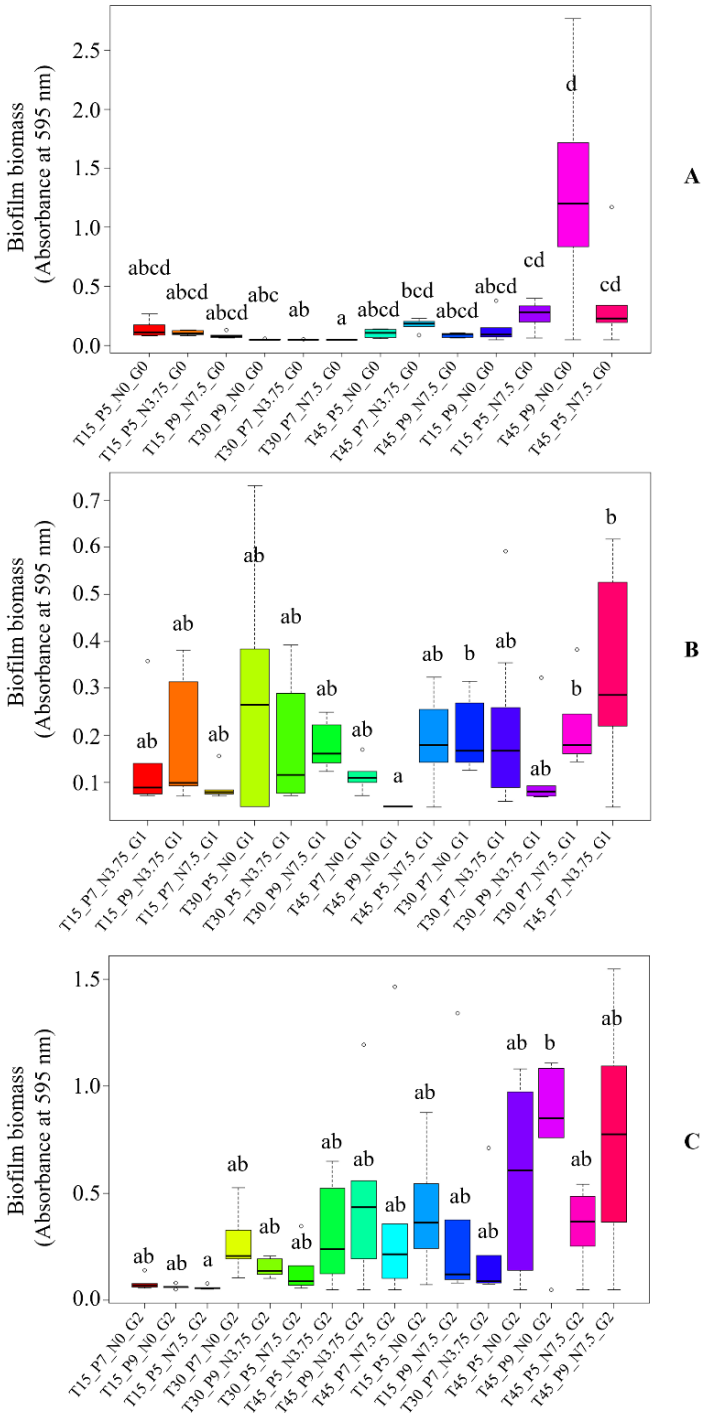

Figure 4: Distribution of absorbances at different glucose concentrations
a) 0%, b) 1%, c) 2% (x-axis = T: Temperature, P: pH, N: NaCl (%), and G: Glucose (%) values)

3.5 Data transformation

To ensure the validity of subsequent analyses, data were first assessed for normality using a combination of visual (histograms and Q-Q plots, as shown in Figure 5) and statistical methods. The Shapiro-Wilk test, specifically, revealed a significant deviation from normality for the raw data (p = 5.72 x 10^{-24}), necessitating the application of a suitable transformation.

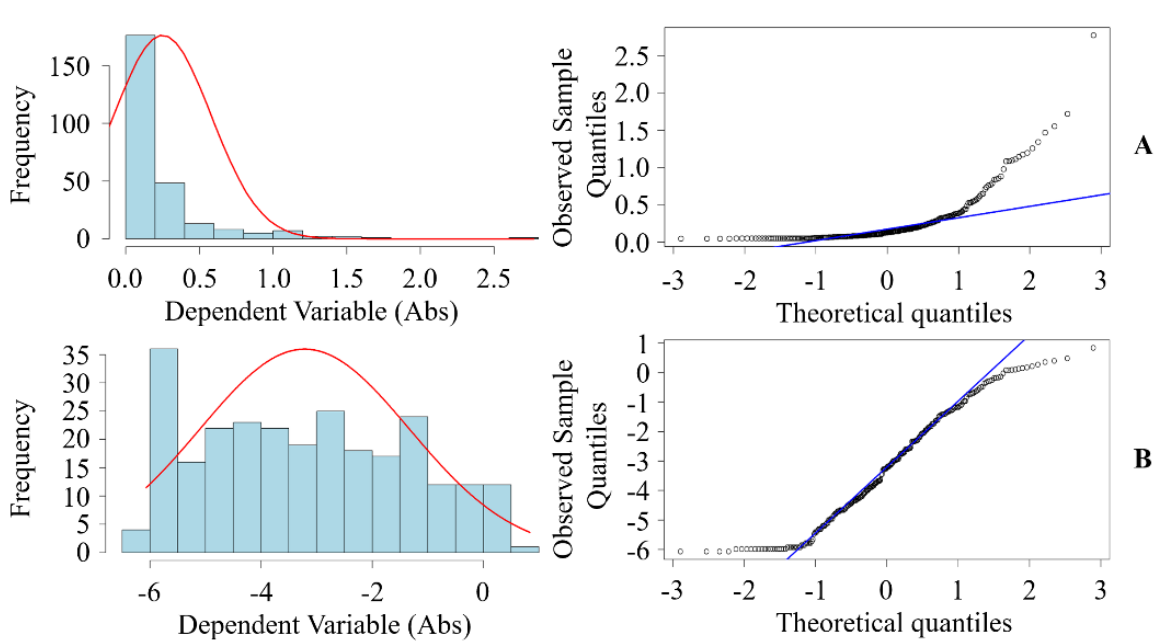

Figure 5: Histograms and Q-Q plots
a) untransformed data, b) Box-Cox transformed data

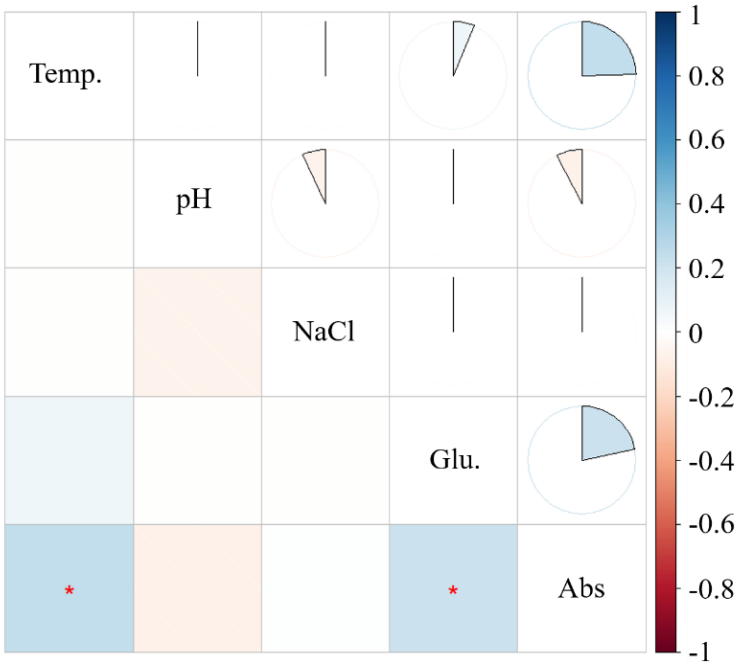

Figure 6: Correlation graph ("*" denotes statistically significant relations) (Temp.: Temperature, Glu.: Glucose concentration, Abs: Box-Cox Transformed Absorbance)

Various transformations, including logarithmic, square root, inverse, exponential, Box-Cox, Yeo-Johnson, and Z-score, were applied and evaluated. The Box-Cox transformation was selected; however, even after the Box-Cox transformation, full adherence to normality or homoscedasticity could not be achieved (p = 4.11 x 10^{-7}). Visual representations of the untransformed and Box-Cox transformed data, including histograms and Q-Q plots, are provided in Figure 5.

3.6 Correlation Analysis

To understand the monotonic relationships among the independent parameters (temperature, pH, glucose, and NaCl concentrations) and their association with the Box-Cox transformed absorbance, a Spearman's rank correlation analysis was conducted. This non-parametric approach was chosen given that the transformed data, while optimised, did not achieve full adherence to normality or homoscedasticity. The results, including their statistical significance, are visually presented in Figure 6.

3.7 Regression Analysis

3.7.1 Permutation-based ANOVA

A statistical test was conducted to investigate the main and interaction effects of the independent parameters on the Box-Cox transformed absorbance. Generally, parametric ANOVA is utilised for such investigations. However, given that the Box-Cox transformed data still did not fully satisfy the parametric assumptions (as discussed in the 'Data transformation' section), a non-parametric alternative, Permutation-based ANOVA, was employed for this purpose.

The results of the Permutation-based ANOVA revealed statistically significant effects on biofilm absorbances. Specifically, a significant effect was observed for glucose concentration (p = 0.0479). Furthermore, highly significant interaction effects were identified: the three-way interactions of pH:Glucose Concentration:NaCl Concentration (p < 0.001) and temperature:pH:NaCl Concentration (p < 0.001) substantially affected biofilm formation. A significant two-way interaction between Temperature and Glucose Concentration (p = 0.0184) was also detected.

3.8 Machine Learning Models

To develop robust predictive frameworks, the normalised dataset was randomly split into training (70%) and test (30%) datasets. Hyperparameters for the machine learning models were optimised using Genetic Algorithm (GA) and Particle Swarm Optimisation (PSO) techniques. The predictive performance of each model was rigorously evaluated using standard metrics including R^2, Adjusted R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), with detailed results presented in Table 1.

The results in Table 1 show that even though the adjusted R^2 value is low, the XGBoost regression model was selected as the best model because it has the highest R^2 and adjusted R^2 values among the models, along with comparatively low RMSE and MAE values.

Table 1: Machine learning models and their predictive performance

Machine Learning Model	R^2	Adjusted R^2	RMSE	MAE
Linear regression	0.2251	0.1780	1.6586	1.3351
Quadratic regression	0.3008	0.2237	1.6626	1.3414
Ridge regression	0.1058	0.0919	1.7817	1.4801
Lasso regression	0.1034	0.0895	1.7840	1.4903
Elastic net regression	0.1043	0.0904	1.7831	1.4876
Polynomial regression	0.4895	0.3920	1.3462	0.9386
Stepwise regression	0.1001	0.0932	1.7873	1.4959
Decision tree regression	0.2956	0.2847	1.5812	1.2383
Random forest regression	0.2042	0.1919	1.5792	1.3136
Gradient boosting	0.1222	0.1086	1.7652	1.4734
XGBoost regression	0.6209	0.6209	1.2355	0.9384
Light GBM regression	0.4136	0.4045	1.4428	1.1022
CatBoost regression	0.4883	0.4804	1.3478	0.9502

Although the adjusted R^2 value (0.6209) indicates that approximately 62% of the variance in biofilm formation was explained by the model, this level of explanatory power is considered acceptable for complex biological systems characterised by multifactorial interactions and inherent experimental variability. Moreover, XGBoost exhibited the best trade-off between predictive accuracy (lowest RMSE and MAE) and generalisation capability among all tested algorithms, justifying its selection as the optimal model for this dataset.

This model was used to analyse the data.

3.8.1 XGBoost regression model

XGBoost (Extreme Gradient Boosting) is a highly efficient and flexible open-source machine-learning algorithm that has gained widespread popularity in various predictive modelling tasks, including regression [22]. It is an ensemble method based on the gradient boosting framework, where new models are iteratively added to correct the errors of previous models. XGBoost is particularly known for its speed, performance, and ability to handle complex datasets with high accuracy while effectively mitigating overfitting through advanced regularisation techniques. Its core architecture relies on decision trees, leveraging parallel processing and tree-pruning strategies to optimise model training.

The algorithm's strength lies in its optimised distributed gradient boosting library, which supports both regression and classification problems. Key features of XGBoost, such as built-in cross-validation, regularisation (L1 and L2), and handling of missing values, contribute to its robustness and superior predictive power, making it an excellent choice for complex predictive modelling like biofilm formation in *K. pneumoniae* [22].

To optimise the performance of the XGBoost model, it was trained with specific hyperparameters. These included 100 boosting iterations (nrounds = 100), a maximum tree depth of 7 (max_depth = 7), and a learning rate (eta) of 0.05. Each individual tree was constructed using all available features (colsample_bytree = 1) to ensure comprehensive learning and help further reduce overfitting. Additionally, the minimal sum of instance weight required in a child node (min_child_weight) was set at 5 to control model complexity.

3.8.2 XGBoost feature importance analysis

The feature importance analysis, derived from the optimal XGBoost regression model for the Box-Cox transformed absorbance, is presented in Figure 7. This analysis quantifies the relative contribution of each environmental parameter to the model's predictive performance.

As depicted in Figure 7, temperature was identified as the most influential factor, exhibiting the highest importance score of 0.33. This indicates that temperature played the most significant role in predicting *K. pneumoniae* biofilm formation. Glucose concentration was followed closely, showing a score of substantial importance of 0.26, highlighting its strong impact on biofilm development.

Conversely, pH demonstrated a moderate but notable contribution, with an importance score of 0.23. NaCl concentration was found to be the least influential factor among the tested parameters, with an importance score of 0.18.

These findings suggest that while all four environmental factors contribute to varying degrees, temperature and glucose concentration are the primary drivers of biofilm formation according to the XGBoost model, with pH and NaCl playing secondary roles in the predictive framework.

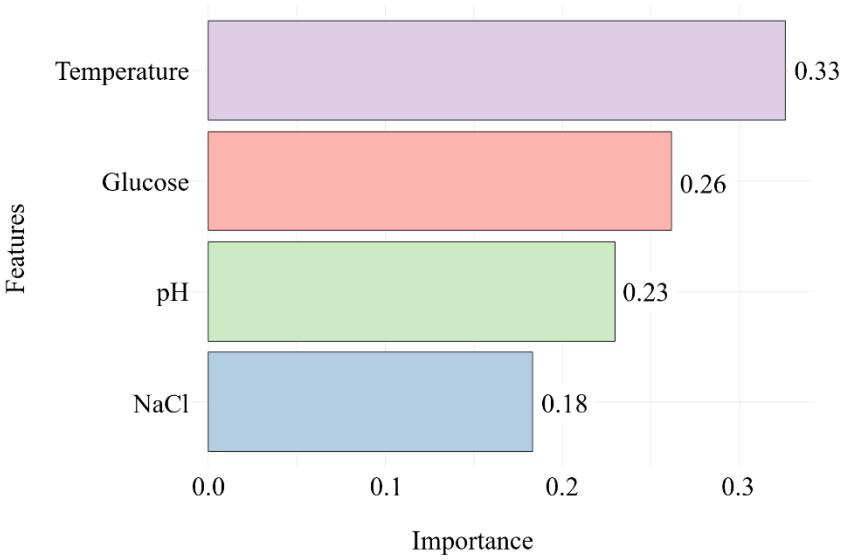

Figure 7: Feature importance plot for the XGBoost - Abs Box-Cox transformed model

3.8.3 Partial dependence analysis for model interpretation

To provide a deeper understanding of how each parameter individually influences the model's predictions, Partial Dependence Plots (PDPs) were generated from the optimal XGBoost model. These plots, presented in Figure 8, illustrate the effect of each feature on the predicted Box-Cox transformed absorbance, after accounting for the average effect of all other features.

The PDP for temperature (Figure 8a) indicates that there is a clear positive and approximately linear relationship; as temperature increases from 15°C to 45°C, the predicted biofilm absorbance consistently rises. This highlights temperature as a factor that directly enhances *K. pneumoniae* biofilm formation.

For glucose concentration (Figure 8b), predicted biofilm absorbance generally increases with a glucose concentration up to 1%, reaching its peak around this point. Beyond 1%, the predicted absorbance tends to decrease or plateau, suggesting an optimal glucose concentration range for biofilm formation.

The pH PDP (Figure 8c) shows that the model predicts the highest biofilm absorbance at neutral pH values (around 7.0). Both more acidic (pH 5.0) and more alkaline (pH 9.0) conditions are associated with significantly lower predicted absorbances, indicating that deviations from neutral pH negatively impact biofilm development.

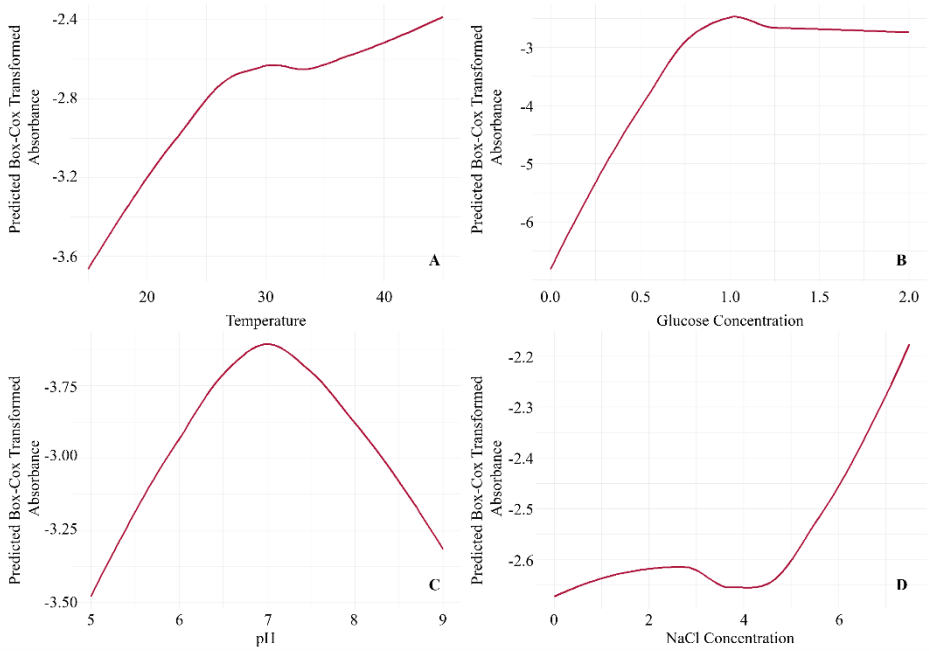

Figure 8: The effect of each factor on absorbance prediction
a) Temperature, b) Glucose concentration, c) pH, and d) NaCl concentration

The analysis for NaCl concentration (Figure 8d) reveals that predicted biofilm absorbance remains relatively stable or slightly increases at lower concentrations (0-3.75%). However, at higher NaCl concentrations, such as 7.5%, there is a sharp increase in predicted absorbance, signifying a strong activating effect of elevated salinity on biofilm formation.

The present study aimed to develop a robust predictive model for *Klebsiella pneumoniae* biofilm formation under varying environmental conditions, specifically focusing on temperature, pH, glucose, and NaCl concentrations. By employing a comprehensive approach that combined experimental design, non-parametric statistical analyses, and advanced machine learning techniques, it was successfully elucidated that the individual and interactive effects of these parameters offer valuable insights into the environmental regulation of *K. pneumoniae* biofilm development.

The findings from the Kruskal-Wallis tests and Permutation-based ANOVA consistently demonstrated that all four investigated factors significantly influenced *K. pneumoniae* biofilm formation. This aligns with extensive previous research highlighting the critical role of physicochemical cues in modulating bacterial biofilm phenotypes across various species [23,24]. The observed significance of complex three-way and two-way interactions, namely pH:Glucose:NaCl, Temperature:pH:NaCl, Temperature:Glucose, underscores that biofilm formation is not merely a sum of individual effects but a dynamic process governed by intricate, synergistic or antagonistic relationships between environmental factors [24]. This complexity further justifies the need for advanced modelling approaches like XGBoost to capture these non-linear interdependencies effectively.

The XGBoost model, identified as the best-performing model in the comprehensive comparison, provided a robust framework for predicting biofilm absorbance. The feature importance analysis clearly identified temperature and glucose concentration as the most influential factors, consistent with their fundamental roles in bacterial physiology and metabolism.

The results demonstrated a clear positive correlation between increasing glucose concentration and predicted biofilm absorbance up to 1% glucose concentration. This corroborates the role of glucose as a key enhancer of biofilm formation. This is in strong agreement with numerous studies showing that readily available carbon sources like glucose promote robust biofilm development in various bacteria, including *K. pneumoniae*, by fueling bacterial growth and extracellular polymeric substance (EPS) synthesis, which is crucial for biofilm matrix integrity [13,24].

Some studies, however, report that excessively high glucose concentrations can sometimes lead to decreased biofilm formation in other species, possibly due to acidic by-product accumulation or osmoregulation challenges, as it was also observed in the present study, where the predicted biofilm absorbance tends to plateau [25-27].

Similarly, temperature played a crucial role in biofilm formation. This is a well-established optimal growth temperature for many mesophilic human pathogens like *K. pneumoniae*, where metabolic efficiency and growth rates are maximised, thus facilitating robust biofilm development [13,28,29].

Müller et al. (2024) also reported that increasing temperatures led to a higher affinity for *K. pneumoniae* strain to adhere to abiotic surfaces, noting a significant increase in specific biofilm formation at 40°C and 42°C compared to 37°C [30]. Regarding biofilm development, temperatures below 37°C, Gual-de-Torrella (2022) proposed that at 37°C, biofilm formation was observed in 76% of *K. pneumoniae* isolates tested, which decreases to 41% 25°C [31]. They also confirmed partially consistent results with previous studies where a temperature increase was favourable for biofilm formation in some *K. pneumoniae* isolates, which supports our results.

The analysis in this present study revealed that pH modulated biofilm formation, with neutral conditions (pH 7.0) promoting optimal biofilm production. Both acidic (pH 5.0) and alkaline (pH 9.0) environments led to reduced predicted biofilm absorbance. This is consistent with the fact that drastic pH changes can denature enzymes, disrupt membrane potential, and impair cellular processes, thereby hindering bacterial growth and adhesion [24,32].

Jastaniah et al. (2022) also identified that pH 7.0 and pH 7.6 resulted in the largest values of exopolysaccharides for *K. pneumoniae*, designating them as the optimum pH range [33].

Monsi et al. observed that biofilm formation was significantly higher in pH 7.2 compared to both the acidic (pH 4.0) and alkaline (pH 11.7) media under both aerobic and anaerobic conditions [34].

The results showed that the biofilm formation sharply increased after 3.75%. Increasing NaCl concentration could enhance biofilm due to several physiological and molecular reasons. Higher salt concentrations create osmotic stress, triggering bacterial stress responses that promote biofilm development as a protective mechanism. This includes increased production of EPS, which stabilises the biofilm matrix and improves cell adhesion.

Murugavel et al. (2024) reported that expression of the LPS gene (WbaG) in *K. pneumoniae* promoted biofilm formation in the presence of Na^+, as well as K^+ and Ca^{2+} [35]. This could be the potential reason for the increase in biofilm in our study. However, Mirkar et al. (2016) found that as salt concentration increased, the biofilm formation of Klebsiella was inhibited at 6% NaCl and above [36]. The reason for this difference could possibly be due to the different strains used in our study and their study.

Despite its strengths, our study has certain limitations. The model was developed using a single *K. pneumoniae* strain (ATCC BAA-1706), and the generalizability of these findings to other clinical isolates or different bacterial species may vary. Future research should therefore aim to validate and expand this model by including a broader range of *K. pneumoniae* strains with diverse genetic backgrounds and by incorporating additional environmental stressors or antimicrobial agents.

In addition, it is primarily focused on biofilm biomass quantification using the crystal violet assay. While this method provides a robust measure of total biofilm, it does not directly assess bacterial viability or metabolic activity within the biofilm. Future studies could incorporate techniques such as CFU counts, metabolic assays, or live/dead staining to provide a more comprehensive characterisation of biofilm formation.

Furthermore, while PDPs provide insights into effects, a deeper exploration of specific feature interactions using more complex visualisation tools could offer even more granular understanding. Integrating real-time monitoring of biofilm parameters with this predictive framework could also enhance its practical utility, allowing for dynamic intervention strategies.

4 Conclusions

The developed XGBoost model demonstrated accuracy in predicting *K. pneumoniae* biofilm biomass. Feature importance analysis revealed temperature and glucose concentration as the most influential factors, significantly impacting biofilm formation. Partial dependence analysis showed that optimal biofilm production was observed at neutral pH and glucose concentrations over 1%, with high temperatures. In addition, high NaCl concentrations significantly stimulated biofilm formation.

These findings provide valuable insights into the regulation of *K. pneumoniae* biofilm formation, which could have practical uses in diverse industrial, medical, and food safety settings. This understanding can inform targeted strategies for controlling biofilm proliferation and enhancing antimicrobial efficacy. Future work could expand this model to other bacterial strains and environmental stressors, integrating it with real-time monitoring and advanced machine learning for even more robust predictive and control capabilities.

In hospital settings, this model could be used to estimate the risk of biofilm formation on medical devices, such as catheters and ventilator components, under specific environmental conditions, thereby supporting preventive strategies, such as environmental adjustments or targeted disinfection protocols. In industrial systems, particularly in water distribution networks or food-processing pipelines, the model may assist in predicting conditions that favour biofilm development and guiding the optimisation of cleaning schedules and salinity or temperature control measures to reduce biofouling.

Furthermore, integration with microfluidic or real-time sensor-based systems could enhance the model's applicability in practical and industrial biofilm control scenarios.

Author Contribution Statements

Conceptualisation, E.M.A.; data curation, N.A. and E.M.A.; methodology, N.A. and E.M.A..; writing, N.A. and E.M.A.; supervision, E.M.A. All authors have read and agreed to the manuscript.

Declaration of Competing Interests

The authors declare no conflict of interest.

Ethical Statement

This research did not involve human participants or animals. Therefore, no ethical approval was required.

Impact Statement

This study provides a data-driven predictive model quantifying the individual and interactive effects of key environmental factors on *Klebsiella pneumoniae* biofilm formation. This model offers a novel, practical tool for identifying conditions that promote or inhibit *K. pneumoniae* biofilm growth, thereby informing targeted strategies for improved infection control in healthcare and enhanced biosecurity in industrial and food safety environments.

Declaration of Generative Artificial Intelligence Use

No artificial intelligence-based tools or applications were used in the preparation of this study. The entire content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

References

[1] Achinas, S., Charalampogiannis, N., Euverink, G.J.W., A brief recap of microbial adhesion and biofilms. *Applied Sciences*, 9 (2019), 2801. https://doi.org/10.3390/app9142801

[2] Donlan, R.M., Biofilms: microbial life on surfaces. *Emerging Infectious Diseases*, 8 (2002), 881-890. https://doi.org/10.3201/eid0809.020063

[3] Du Toit, A., Bacterial architects build the biofilm structures. *Nature Reviews Microbiology* (2024), 187. https://doi.org/10.1038/s41579-024-01020-6

[4] Dufrêne, Y.F., Viljoen, A., Binding strength of gram-positive bacterial adhesins. *Frontiers in Microbiology*, 11 (2020), 554551. https://doi.org/10.3389/fmicb.2020.01457

[5] Feng, X., Wu, Q., Che, L., Ren, N., Analysing the inhibitory effect of metabolic uncoupler on bacterial initial attachment and biofilm development and the underlying mechanism. *Environmental Research*, 185 (2020), 109390. https://doi.org/10.1016/j.envres.2020.109390

[6] Joseph, R.L., Prosthetic joint infections: Bane of orthopedists, challenge for infectious disease specialists. *Clinical Infectious Diseases*, 36 (2004), 1157-1161. https://doi.org/10.1086/374554

[7] Roy, R., Tiwari, M., Donelli, G., Tiwari, V., Strategies for combating bacterial biofilms: A focus on anti-biofilm agents and their mechanisms of action. *Virulence*, 9 (2018), 522-554. https://doi.org/10.1080/21505594.2017.1313372

[8] Samrot, A.V., Abubakar Mohamed, A., Faradjeva, E., Si Jie, L., Hooi Sze, C., Arif, A., Sean, T.C., Michael, E.N., Mun, C. Y., Qi, N.X., Mok, P.L., Kumar, S.S., Mechanisms and impact of biofilms and targeting of biofilms using bioactive compounds: A review. *Medicina*, 57 (2021), 839. https://doi.org/10.3390/medicina57080839

[9] Markou, G., Georgakakis, D., Cultivation of filamentous cyanobacteria (blue-green algae) in agro-industrial wastes and wastewaters: a review. *Applied Energy*, 88 (2011), 3389-3401. https://doi.org/10.1016/j.apenergy.2010.12.042

[10] Pan, M., Lyu, T., Zhang, M., Zhang, H., Bi, L., Wang, L., Chen, J., Yao, C., Ali, J., Best, S., Ray, N., Pan, G., Synergistic recapturing of external and internal phosphorus for *in situ* eutrophication mitigation. *Water*, 12 (2019), 1-9. https://doi.org/10.3390/w12010002

[11] Liu, H. Y., Prentice, E. L., Webber, M. A., Mechanisms of antimicrobial resistance in biofilms. *NPJ Antimicrobials and Resistance*, 2 (1) (2024), 27. https://doi.org/10.1038/s44259-024-00046-3

[12] Li, L., Gao, X., Li, M., Liu, Y., Ma, J., Wang, X., Yu, Z., Cheng, W., Zhang, W., Sun, H., Song, X., Relationship between biofilm formation and antibiotic resistance of Klebsiella pneumoniae and updates on antibiofilm therapeutic strategies. *Frontiers in Cellular and Infection Microbiology*, 14 (2024), 1324895. https://doi.org/10.3389/fcimb.2024.1324895

[13] Guerra, M.E.S., Destro, G., Vieira, B., Lima, A.S., Ferraz, L.F.C., Hakansson, A.P., Darrieux, M., Converso, T.R., Klebsiella pneumoniae biofilms and their role in disease pathogenesis. *Frontiers in Cellular and Infection Microbiology*, 12 (2022), 877995. https://doi.org/10.3389/fcimb.2022.877995

[14] Khalefa, H.S., Arafa, A.A., Hamza, D., El-Razik, K.A.A., Ahmed, Z., Emerging biofilm formation and disinfectant susceptibility of ESBL-producing Klebsiella pneumoniae. *Scientific Reports*, 15 (1) (2025), 1599. https://doi.org/10.1038/s41598-024-84149-x

[15] Rahdar, H.A., Malekabad, E.S., Dadashi, A.R., Takei, E., Keikha, M., Kazemian, H., Karami-Zarandi, M., Correlation between biofilm formation and carbapenem resistance among clinical isolates of Klebsiella pneumoniae. *Ethiopian Journal of Health Sciences*, 29 (6) (2019). https://doi.org/10.4314/ejhs.v29i6.11

[16] Al-Bayati, M., Samarasinghe, S., Biofilm and gene expression characteristics of the carbapenem-resistant enterobacterales, Escherichia coli IMP, and Klebsiella pneumoniae NDM-1 associated with common bacterial infections. *International Journal of Environmental Research and Public Health*, 19 (8) (2022), 4788. https://doi.org/10.3390/ijerph19084788

[17] Sooriyakumar, P., Bolan, N., Kumar, M., Singh, L., Yu, Y., Li, Y., Weralupitiya, C., Vithanage, M., Ramanayaka, S., Sarkar, B., Wang, F., Gleeson, D.B., Zhang, D., Kirkham, M.B., Rinklebe, J., Siddique, K.H.M., Biofilm formation and its implications on the properties and fate of microplastics in aquatic environments: a review. *Journal of Hazardous Materials Advances*, 6 (2022), 100077. https://doi.org/10.1016/j.hazadv.2022.100077

[18] Pan, Y., Breidt Jr, F., Gorski, L., Synergistic effects of sodium chloride, glucose, and temperature on biofilm formation by Listeria monocytogenes serotype 1/2a and 4b strains. *Applied and Environmental Microbiology*, 76 (5) (2010), 1433-1441. https://doi.org/10.1128/AEM.02185-09

[19] Silva, V., Pereira, J.E., Maltez, L., Poeta, P., Igrejas, G., Influence of environmental factors on biofilm formation of Staphylococci isolated from wastewater and surface water. *Pathogens*, 11 (10) (2022), 1069. https://doi.org/10.3390/pathogens11101069

[20] Zurnacı, M., Şenturan, M., Şener, N., Gür, M., Altınöz, E., Şener, İ., Altuner, E.M., Studies on antimicrobial, antibiofilm, efflux pump inhibiting, and ADMET properties of newly synthesised 1, 3, 4-thiadiazole derivatives. *ChemistrySelect*, 6 (45) (2021), 12571-12581. https://doi.org/10.1002/slct.202103214

[21] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. (accessed 2025).

[22] Chen, T., Guestrin, C., Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, (2016), 785-794. https://doi.org/10.1145/2939672.2939785

[23] Singh, A.K., Yadav, S., Chauhan, B.S., Nandy, N., Singh, R., Neogi, K., Roy, J.K., Srikrishna, S., Singh, R.K., Prakash, P., Classification of clinical isolates of Klebsiella pneumoniae based on their in vitro biofilm forming capabilities and elucidation of the biofilm matrix chemistry with special reference to the protein content. *Frontiers in Microbiology*, 10 (2019), 669. https://doi.org/10.3389/fmicb.2019.00669

[24] Li, Y., Ni, M., Regulation of biofilm formation in Klebsiella pneumoniae. *Frontiers in Microbiology*, 14 (2023), 1238482. https://doi.org/10.3389/fmicb.2023.1238482

[25] Chen, T., Xu, Y., Xu, W., Liao, W., Xu, C., Zhang, X., Cao, J., Zhou, T., Hypertonic glucose inhibits growth and attenuates virulence factors of multidrug-resistant Pseudomonas aeruginosa. *BMC Microbiology*, 20 (1) (2020), 203. https://doi.org/10.1186/s12866-020-01889-2

[26] Vashchenko, A.O., Voronkova, Y.S., Kulyk, E.E., Snisar, O.S., Sidashenko, O.I., Voronkova, O.S., Influence of sugars on biofilm formation of Staphylococcus epidermidis. *Regulatory Mechanisms in Biosystems*, 12 (2) (2021), 321-325. https://doi.org/10.15421/022143

[27] Mohammed, S., Hadi, H.M., Hamed, N.A.H.K., Abbas, N.K., The effect of Different concentrations of glucose solution on the growth and biofilm formation of sensitive and resistant Pseudomonas aeroginosa isolates. *Kerbala Journal of Pure Sciences*, 19 (2) (2022), 140-148.

[28] Ramakrishnan, R., Nair, A.V., Parmar, K., Rajmani, R.S., Chakravortty, D., Das, D., Combating biofilm-associated Klebsiella pneumoniae infections using a bovine microbial enzyme. *NPJ Biofilms and Microbiomes*, 10 (1) (2024), 119. https://doi.org/10.1038/s41522-024-00593-7

[29] Beckman, I.V., R.L., Cella, E., Azarian, T., Rendueles, O., Fleeman, R.M., Diverse polysaccharide production and biofilm formation abilities of clinical Klebsiella pneumoniae. *NPJ Biofilms and Microbiomes*, 10 (1) (2024), 151. https://doi.org/10.1038/s41522-024-00629-y

[30] Müller, J.U., Schwabe, M., Swiatek, L.S., Heiden, S.E., Schlüter, R., Sittner, M., Bohnert, J.A., Becker, K., Idelevich, E.A., Guenther, S., Schaufler, K., Temperatures above 37° C increase virulence of a convergent Klebsiella pneumoniae sequence type 307 strain. *Frontiers in Cellular and Infection Microbiology*, 14 (2024), 1411286. https://doi.org/10.3389/fcimb.2024.1411286

[31] Gual-de-Torrella, A., Delgado-Valverde, M., Perez-Palacios, P., Oteo-Iglesias, J., Rojo-Molinero, E., Macià, M.D., Oliver, A., Pascual, A., Fernandez-Cuenca, F., Prevalence of the fimbrial operon mrkABCD, mrkA expression, biofilm formation and effect of biocides on biofilm formation in carbapenemase-producing Klebsiella pneumoniae isolates belonging or not belonging to high-risk clones. *International Journal of Antimicrobial Agents*, 60 (4) (2022), 106663. https://doi.org/10.1016/j.ijantimicag.2022.106663

[32] Herrera-Espejo, S., Domínguez-Miranda, J.L., Rodríguez-Mogollo, J.I., Pachón, J., Cordero, E., Pachón-Ibáñez, M.E., Effects of pH on the pathogenicity of Escherichia coli and Klebsiella pneumoniae on the kidney: In vitro and in vivo studies. *International Journal of Molecular Sciences*, 25 (14) (2024), 7925. https://doi.org/10.3390/ijms25147925

[33] Jastaniah, S.D., Jamal, T.Y., Amashah, R.H., Aly, M.M., New strategies for inhibition of Listeria monocytogenes and Klebsiella pneumoniae biofilm formation and persistence. *Journal of Contemporary Medical Sciences*, 8 (6) (2022), 400-407. https://doi.org/10.22317/jcms.v8i6.1302

[34] Monsi, T.P., Giami, L.K., Ganabel, C.B., Modulation of virulent factors in Klebsiella pneumoniae exposed to different hydrogen ion concentrations. *Microbiology Archives, an International Journal*, 6 (2) (2024), 27-32. https://doi.org/10.51470/MA.2024.6.2.27

[35] Murugavel, A., Gunasekaran, S., Ramakrishnan, J., Elucidating the effect of cations on Klebsiella pneumoniae biofilm and related genes. *bioRxiv*, (2024), 2024-01. https://doi.org/10.1101/2024.01.24.577145

[36] Mirkar, K., Rawat, A., Satish, R., Effect of environmental factors on biofilm formation. *Indian Journal of Life Sciences*, 5 (2) (2016), 53-64.