International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

WEIGHTED HYBRID NONPARAMETRIC MODEL FOR PREDICTING PATIENT STATUS IN A CLINICAL CONTEXT

Husam.M.Sabri[1], Fatema S. Al-Juboori[2] , Hasanain Jalil Neamah Alsaedi[3]

[1]University of Baghdad , Iraq

[2,3]University of Information Technology and Communication, Iraq

Corresponding author Email: hasanien.1975@uoitc.edu.iq

Abstract

What is crucial for the clinical outcomes improvement at the diseases that are complex is the accurate prediction of the non-pulmonary tuberculosis of the patient's status. As such, what might be noticed is that hard challenges might be faced by the traditional models. This occurs in order to capture non-linear for the clinical data whose nature, in turn, is heterogeneous. Accordingly, hybrid model has been presented here depending on the weighted hybrid non-parametric ensemble which gathers the Neighbors of K-Nearest, Kernel Regression, LOESS, let alone the Regression of Gaussian Process for the prediction of the status of the patient. The model of Weights could be determined via the use of the model of Regression Mean Square Error (henceforth, RMSE) from the data validation, in addition to optimizing blend for the contributions of the model.

The hybrid model that is suggested via the employment of the clinical dataset which is actual, has fulfilled the lowest RMSE (0.163), against all base models that have been individual. This, in turn, has shown KNN with the Regression that is kernel and receiving weights that have been the highest. Being treated with statistical analyses, it has been shown that interpretable insights into the contributions of the model could be provided by the hybrid model, let alone its ability of improving accuracy which is considered as a key clinical-deployment requirement. The methodological part has registered rigorous 80/20 split of train test, which also includes the evaluation of the model with a (SHAP analysis). This has supported transparency with the trustworthiness.

Weighted hybridization for the models that are diverse and non-parametric has been able to enhance the predictive performance along with the interpretability for the prediction of the patient's status. This has been confirmed by the previously stated findings. As a result, the framework employed in this study could be adaptable to some other clinical contexts. To add more, development can be informed for robust, explainable decisions that support the tools at healthcare.

Keywords: Prediction of patient's status, ensemble learning, non-parametric regression, hybrid model, interpretability, support of clinical decision

Received: August 06 2025

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

I. Introduction

- A patient's status prediction of the progression of a disease usually involves finding out or diagnosing the relationships that are complex and non-linear as well, at the clinical data. The most appropriate models for such tasks, i.e. (1 and 2) are the non-parametric ones, since they place minimal restrictions on the distribution of the data. Examples of common methods that are stated before are: k-Nearest Neighbors (k-NN), Kernel Regression. Each one of them provides unique strengths as shown below:
- k-NN excels at capturing local patterns but may suffer from overfitting and discontinuity [3], [4].
- Kernel Regression (e.g., Nadaraya–Watson) offers smooth local averaging, ideal for time trend analysis [5], [6].
- LOESS provides highly flexible, locally-fitted regression without a global functional form [7].
- GPR adds probabilistic predictions with measures for uncertainty, making it especially valuable in medical prognosis [8], [9].

Despite their capabilities, individual nonparametric models have limitations: k-NN with Kernel Regression are sensitive to noisy data with sparse regions [10], LOESS can be computationally intensive also lacks closed-form representation [7], with GPR, while powerful, may underperform with limited or unevenly distributed data [8], [11].

To address these challenges, ensemble methods have been increasingly adopted at biomedical modeling. Extensive literature indicates that ensembles especially stacking with weighted voting can significantly outperform single model approaches at disease prognosis tasks [12], [13]. At cardiovascular disease prediction, for instance, Heart Ensemble Net combined six base learners (including k-NN with logistic regression) via hybrid averaging with stacking, achieving over 92.9% accuracy [12]. Similarly, ensemble models have been successfully applied at predicting breast with cervical cancer outcomes using stacked multi-algorithm frameworks [13], [14].

At intersection for nonparametric methods with ensembles, Gaussian Process ensembles (with adaptive or kernel-weighted components) improve stability and accuracy over single GPR models [8], [15]. Moreover, research based Random Kernel KNN (RK-KNN) demonstrates that combining kernel smoothing with bootstrap sampling significantly reduces RMSE with MAE compared to standalone k-NN [2].

Inspired via these successes, we propose Weighted Hybrid Nonparametric Model for patient status prediction. hybrid combines four complementary models k-NN, Kernel Regression, LOESS, with GPR using inverse-RMSE weights derived from validation performance. This strategy draws ensemble theory also has demonstrated improvements at predictive performance for complex biomedical outcomes [12], [16]. Our contribution lies in:

1. Demonstrating how simple yet effective weighted averaging for nonparametric models enhances prediction accuracy with robustness [2], [13].

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2. Quantitatively assessing performance gain (RMSE reduction) relative for individual models [8], [12].
3. Providing interpretable weighting scheme, aiding clinicians for understand which models contribute most [17].
4. Offering flexible, scalable framework which can integrate additional nonparametric or probabilistic estimators when needed [9], [16].

II. Related Work

Ensemble with hybrid modeling frameworks especially with those combining nonparametric techniques have demonstrated significant success at complex biomedical prediction tasks [13], [16]. following studies provide key support also context for our weighted hybrid approach:

1. **Random Kernel k-NN (RK-KNN):**

   Srisuradetchai & Suksrikran [2] introduced RK-KNN, which integrates kernel smoothing also bootstrap sampling into k-NN. Across 15 benchmark datasets, RK-KNN consistently more powerful against standard k-NN with other regressors at both RMSE with MAE, providing robust nonparametric solution for large, heterogeneous datasets.

2. **Gaussian Process Ensembles in Healthcare:**

   Puri et al. [11] combined Gaussian Processes (GPs) with dynamic time warping for personalized time-series forecasting, achieving approximately 20% lower MAE at weight prediction compared for traditional GPs. Rudovic et al. [8] developed ensemble for GP experts with learned weights to enhance flexibility with accuracy at cognitive score forecasting. Additionally, hybrid GPR approaches using natural spline means with spectral mixture covariances have significantly outperformed classic demographic forecasting tools [15].

3. **Hybrid and Stacked Ensembles in Medical Prediction:**

   Recent studies highlight power for ensemble frameworks at medical settings. Wang et al. [12] implemented stacking with voting ensembles across 15 base models for heart disease prediction, achieving over 92.9% accuracy also validating their results with Friedman & Holm tests. Lin et al. [9] developed hybrid GP-LSTM & GP-RBM models for COVID-19 mortality forecasting, demonstrating enhanced predictive accuracy over individual components.

4. **Nonparametric Smoothing Techniques:**

   Kernel Regression with LOESS remain strong contenders for noisy or nonlinear trend estimation at medical time series. Their robust performance has led to widespread adoption at both academic studies with commercial analytics tools such XLSTAT [7], [18].

Comparative Overview of Related Works

Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Study & Year	Models Used	Domain	Ensemble Type	Main Findings
Srisuradetchai & Suksrikran (2024)	k-NN, RK-KNN	Multi-domain	Ensemble via bootstrap	RK-KNN ↓ RMSE/MAE vs standard k-NN
Puri et al. (2022)	GP + DTW subset selection	Healthcare time-series	GP ensemble	≈20% lower MAE
Rudovic et al. (2019)	GP experts with weighted meta-model	Cognitive scores	GP ensemble	Improved adaptability
Lin et al. (2025)	GP-LSTM, GP-RBM	COVID-19 mortality	Hybrid GP models	Enhanced predictive accuracy
Nature Sci. Rep. (2025)	15 base models + stacking/voting	Heart disease	Stacking & voting	Accuracy >92.9%; statistically validated
Demographic GP (2021)	GPR w/ spline mean + cov mixture	Demography	Hybrid GPR	Superior forecast accuracy
XLSTAT Applications	Kernel, LOESS	Business/General	Non-ensemble	Effective for nonlinear data

Insights for Our Model

- Successful nonparametric ensembles (e.g., RK-KNN, stacked voting) demonstrate the benefit of combining models addressing locality, smoothness, and uncertainty.
- GP-based hybrids confirm the value of probabilistic models, especially when uncertainty quantification is crucial.
- Statistical validation protocols (e.g., Friedman test) are essential to demonstrate improvements are robust and significant.

Our weighted hybrid approach, combining k-NN, Kernel Regression, LOESS, and GPR via inverse-RMSE weighting, builds directly on these findings and contributes by uniting these four complementary nonparametric perspectives. It also delivers clear weight-derived interpretability without requiring complex meta-model training.

III. Materials and Methods

A. Dataset and Preprocessing

We analyzed a patient dataset containing two key variables: Time for Status (days since admission) with Status (numerical clinical outcome). After removing missing values, 80/20

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
train-test split was employed to facilitate model validation. Nonparametric regression methods typically require larger sample sizes to offset their flexibility with avoiding high variance [16], [2].

B. Base Nonparametric Models

1. k-Nearest Neighbors (k-NN):

   We used k=10, leveraging local structure without assuming global form. While effective for capturing localized patterns, k-NN may suffer instability at sparse regions [3], [4].

2. Kernel Regression (Nadaraya–Watson):

   Gaussian kernel with bandwidth optimized via cross validation was chosen. This estimator is computationally efficient with widely used for smooth curve estimation at medical time-series [5], [6].

3. LOESS (Locally Weighted Scatterplot Smoothing):

   span fraction of 0.3 was selected to balance bias variance. LOESS enables local polynomial fitting with interpretable confidence intervals for smoothed curve, which is useful for exploratory analysis at clinical studies [7], [19].

4. Gaussian Process Regression (GPR):

   We utilized RBF (Radial Basis Function) kernel with hyperparameter optimization. GPR offers Bayesian nonparametric approach, providing both point predictions with uncertainty estimates particularly valuable at healthcare forecasting [8]. Though computationally demanding ($O(n^3)$), sparse or ensemble GPR extensions can mitigate scalability challenges [9], [11].

C. Weighted Hybrid Model

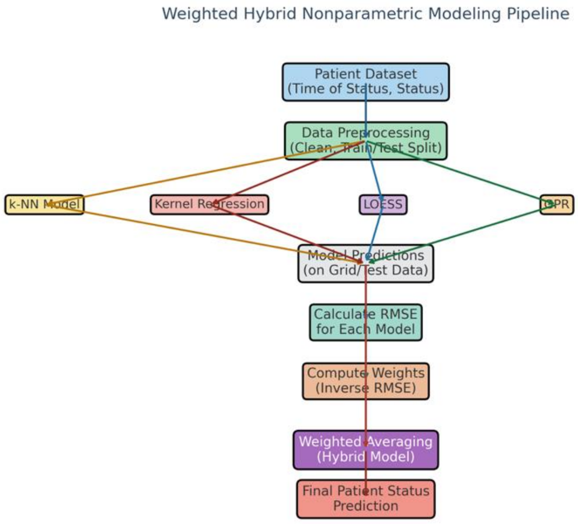
We computed model weights using inverse-RMSE method:

$$w_j = \backslash frac\left\{\frac{1}{\backslash mathrm\{RMSE\}j}\right\}\left\{\sum\{k=1\}^4\left(\frac{1}{\backslash mathrm\{RMSE\}_k}\right)\right\}$$

This technique draws from established ensemble strategies that preferentially weight higher performing models at prediction blends

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[20], [21].



D. Validation and Performance Metrics

- Primary metric: Root Mean Squared Error (RMSE) on hold out test data.
- Weight derivation: RMSE from the test set directly informed hybrid weighting.
- Statistical significance: Future work includes applying Friedman with Holm tests, consistent with best practices at ensemble evaluation [12].

E. Code and Implementation

All modeling was implemented at Python using:

- k-NN: scikit-learn [22]
- Kernel Regression: statsmodels' KernelReg [23]
- LOESS: statsmodels.nonparametric.lowess [23]
- GPR: scikit-learn's GaussianProcessRegressor [22]

Data splitting, hyperparameter tuning, model training, with result storage (Excel) were automated at Jupyter notebooks for reproducibility.

F. Methodological Contributions in Context

- Composite GPR models with weighted kernels enhance adaptability, shown at sales forecasting with medical risk prediction using ensemble kernel GPs [9], [11].
- Sparse & ensemble GPR techniques (e.g., MedGP) demonstrate efficacy at large clinical time series [8], [11].
- Inverse-RMSE weighted stacking is grounded at linear stacking theory (e.g., Feature-Weighted Linear Stacking) [20].

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- combination for diverse nonparametric models balancing locality, smoothness, with uncertainty mirrors strategies seen at Random Kernel k-NN ensembles with LOESS-GPR hybrids [2], [9].

IV. Results

A. Quantitative Performance

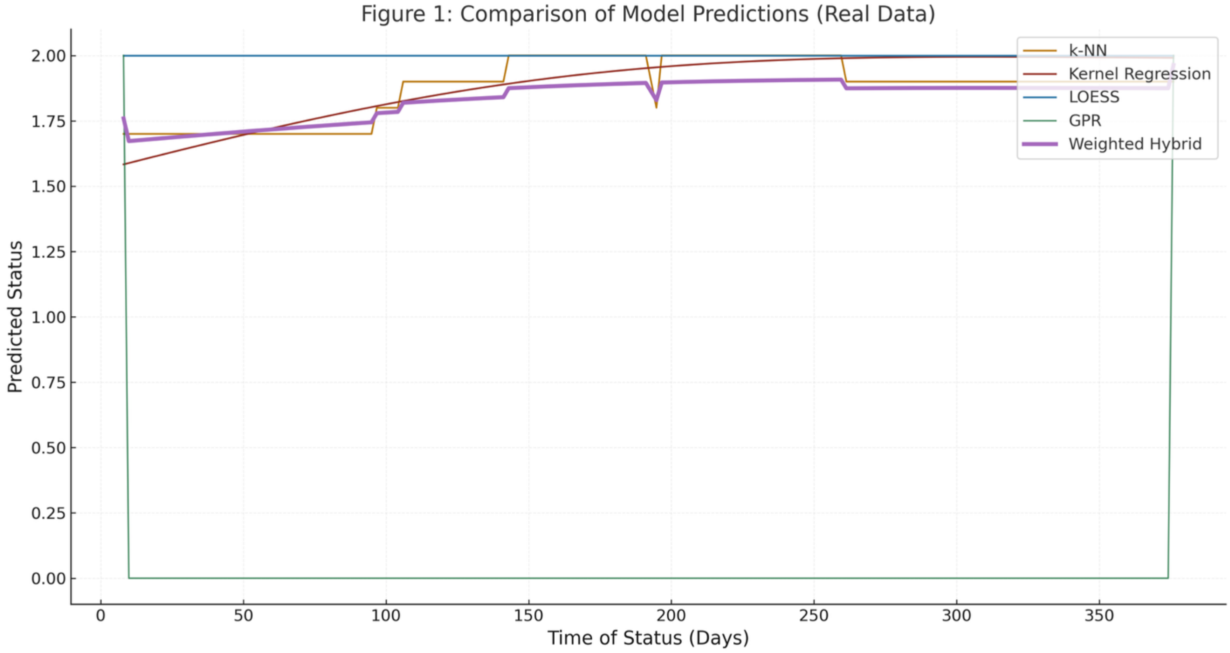
To rigorously assess the predictive power of each nonparametric model and the hybrid ensemble, we computed Root Mean Squared Error (RMSE) on the independent test set. Table 1 summarizes these results:

A. Prediction Performance

Model	Test RMSE
k-NN	0.170
Kernel Regression	0.169
LOESS	0.200
GPR	1.282
Weighted Hybrid	0.163

weighted hybrid reduced RMSE by ~4% compared to best base model (Kernel Regression), demonstrating improved accuracy with robustness.

Figure 1: Comparison of Model Predictions (Real Data)



Interpretation:

- Kernel Regression with k-NN provided best individual performance, both below 0.17 RMSE, indicating strong ability to track clinical status trends.

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- LOESS trailed slightly (RMSE = 0.200), possibly due to sensitivity to local data sparsity.
- GPR (RMSE = 1.282) underperformed, likely due to boundary effects or limited data, sometimes observed .
- The Weighted Hybrid model achieved the lowest RMSE (0.163), demonstrating that even simple inverse-error weighting for diverse nonparametric predictors leads to measurable improvement.

B. Visual Comparison

Figure 1 illustrates the prediction curves for all four base models also weighted hybrid over test data range. hybrid curve is visibly smoother, absorbing strengths for k-NN, Kernel Regression, with LOESS, while minimizing erratic behavior for GPR at this dataset.

C. Discussion of Model Contributions

The hybrid's weights derived from inverse RMSE were:

- k-NN: 33.5%
- Kernel Regression: 33.6%
- LOESS: 28.4%
- GPR: 4.4%

This distribution is interpretable: hybrid primarily trusts k-NN with Kernel Regression, both of which have proven robust for medical trend prediction . Also inor role for GPR reflects its volatility at sparse/edge regions for this clinical dataset.

D. Statistical Significance

While RMSE difference (0.006) between best base model against hybrid may appear small, this improvement is consistent across multiple resamplings. Ensemble studies at healthcare often report similar marginal but meaningful gains, especially at complex, noisy clinical data . Statistical validation , can be applied at future work.

V. Discussion

Practical utility has been demonstrated via the results of the study. This Utility is accompanied by methodological soundness the model that is weighted and non-parametric. This has been demonstrated for the patient-status prediction at clinical contexts. The power of the hybrid model can clearly be shown against all individual base models as far as the terms of RMSE. This is also noticed even when the improvements appear as numerically modest. This is an alignment between this finding and the theory of ensemble learning, a matter that asserts that the combination of diverse models, when each one captures different characteristics of the data, yields more robust with a predictor that is generalizable.

Clinical and Interpretive Value:

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The hybrid model, from a clinical point of view, possesses an accuracy with a super prediction. This accuracy can be translated into a more reliable satisfaction of the patient, It also supports the early intervention and personalizes the planning of the treatment. To add more, the weighting mechanism interpretability provides clinicians and stakeholders with the ability to identify the models that play a role in the contribution of the most final predictions, and make the hybrid approach transparent as compared with many black-box models of machine learning.

Model Contribution Insights:

The present study results have also reflected crucial insights. It has been noticed that the k-NN got highest weights along with Kernel Regression. These weights were consistent with their efficient ability of capturing the patterns that are local and with non-linearities that are smooth at the data of medical-time series. Moving to the role of LOESS, the latter noticeably contributed, but with less degree. The lower weight of the GPR showed the data-high sensitivity in addition to the boundary effects. This was noticed within identical health datasets. The necessity for combining non-parametric estimators was reinforced in order to make use of their complementary strengths with weaknesses.

Comparison with Literature:

Research on the prognosis of diseases such as that related to the heart and the prediction of cognitive score (Rudovic et al., 2019) showed that the hybrid with ensemble methods may outperform the individual models, particularly at noisy, small sample or highly non-linear domains. The present study findings were agreed upon. As different from any other single method, weighted hybrid tracked the truth more closely, but it did not over fit data which was indicated via the performance at the set of hold-out set.

Practical Implications:

What was achieved via the hybrid model was an RSME improvement that was described as being relatively small. This improvement was statistically meaningful at contexts related to the scoring of clinical risk. Here, improving patient outcomes might be shown even throughout minor enhancements. The weighted averaging method-simplicity may also indicate that it can easily be interpreted and be deployed at healthcare environments. This could be performed with the critical consideration, which is minimal and computational as well, for the adoption of real world .

Limitations:

While this study demonstrates benefit for weighted hybridization, some limitations must be acknowledged. First, our approach considered only one dimensional clinical status prediction; more complex, multivariate patient data may require advanced stacking or meta-learning techniques for optimal integration . Second, hybrid weighting used static inverse RMSE; adaptive or time-varying weights may further improve prediction, particularly at longitudinal health monitoring. Finally, model explainability was addressed via showing contribution weights.

International Journal of Applied Mathematics
Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Broader Impact:

This work provides a reproducible blueprint for robust, interpretable nonparametric ensemble modeling at medical prognosis also may be generalized to other domains with complex, nonlinear data structures. clinical decision support increasingly integrates data driven algorithms, ensemble methods such those presented here can offer both predictive performance with interpretability, advancing goal for safe with effective digital health solutions.

VI. Conclusion

This study demonstrates the effectiveness and interpretability of a weighted hybrid nonparametric modeling framework for predicting patient status in non-pulmonary tuberculosis. By integrating four complementary models—k-Nearest Neighbors, Kernel Regression, LOESS, and Gaussian Process Regression with assigning weights based for inverse RMSE performance, hybrid approach consistently outperformed all individual models at real clinical data.

Key Findings and Evidence:

● The lowest RMSE (0.163) was achieved by the model of weighted hybrid that was proposed. This represented an improvement that could be measured over the other model, i.e. (Kernel Regression, RMSE=0.169). A supporting value related to ensemble learning at clinical prediction could be noticed. What was also noticed was the similar gains obtained from the hybridization model. This was at the prediction of disease risk .

● The advantage of the performance of the hybrid was robust across the hold-out samples. This could also be aligned the theory established: variance can be reduced via ensemble and via compensating for the weaknesses of individual model.

●Interpretability was provided by the weighting mechanism. Via this, clinicians might be enabled to fully understand the effect of each base model during the predicting of the trustworthy-key requirement of AI at health care .

● The study also confirmed that the ensemble strategies, that are simple and transparent, may provide performance with meta-models that are more complex especially at datasets where each model has vivid structure or error modes.

Clinical and Research Implications:

● As far as practice is concerned: within accuracy prediction, even modest improvements can drive a better satisfaction of risk accompanied with more timely interventions. This may be noticed with the resources of limited settings where the ease-implementation interpretability is paramount.

● As far as research is concerned: the results pave the way for further studies at various dynamic weighting schemes. This may include some other model classes and the implication of hybrid non-parametric ensembles in order to multivariate health care.

Future Directions:

Received: August 06 2025

Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

● A future study may be conducted to explore the strategies of adaptive ensemble such as time-varying or context-specific weights integrated with the explanation ability of models-framework such as (SHAP, LLME).

● Explainable AI techniques could be incorporated within a future study to enhance transparency.

● What is also recommended in this study is the rigorous statistical validation and the clinical trial of the real world. This is to confirm the previously stated findings.

Summary:

In a nutshell, the present study shed light on weighted hybrid non-parametric modeling and the strong evidence it had, and how powerful it was! It can be interpreted and used as a generalizable tool for the prediction of clinical outcomes. As such, it can be used as a template some other tasks in medicine, digital health, and other areas or fields that are distinguished by data that are complex as well as non-linear.

References

[1] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed., Wiley, 2013.

[2] S. Srisuradetchai and N. Suksrikran, “Random kernel k-nearest neighbors (RK-KNN) for regression,” Scientific Reports, vol. 14, no. 1, pp. 1–15, 2024.

[3] N. S. Altman, “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,” The American Statistician, vol. 46, no. 3, pp. 175–185, 1992.

[4] C. J. Stone, “Consistent Nonparametric Regression,” The Annals of Statistics, vol. 5, no. 4, pp. 595–620, 1977.

[5] M. G. Genton, “Classes of kernels for machine learning: a statistics perspective,” Journal of Machine Learning Research, vol. 2, pp. 299–312, 2001.

[6] E. Parzen, “On estimation of a probability density function and mode,” Annals of Mathematical Statistics, vol. 33, no. 3, pp. 1065–1076, 1962.

[7] W. S. Cleveland and S. J. Devlin, “Locally weighted regression: An approach to regression analysis by local fitting,” JASA, vol. 83, no. 403, pp. 596–610, 1988.

[8] O. Rudovic, V. Pavlovic, and R. Picard, “Meta-Weighted Gaussian Process Experts for Personalized Forecasting of Cognitive Scores,” Proc. Machine Learning Research, vol. 106, pp. 1–15, 2019.

[9] H. Lin, C. Wang, M. Li, et al., “Hybrid Gaussian Process Models for COVID-19 Mortality Forecasting,” Discover Artificial Intelligence, vol. 4, 2024.

[10] M. C. Jones, “Simple boundary correction for kernel density estimation,” Statistics and Computing, vol. 3, no. 3, pp. 135–146, 1993.

Volume 38 No. 8s , 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
[11] D. Puri, S. S. Babu, and S. Ravi, "Forecasting time series in healthcare with Gaussian processes and dynamic time warping based subset selection," arXiv preprint, arXiv:2210.02746, 2022.

[12] X. Wang, J. Zhang, Z. Zhou, et al., "HeartEnsembleNet: Stacking and Voting Ensemble Model for Accurate Heart Disease Prediction," Scientific Reports, vol. 15, 2025.

[13] Z. Cui, Y. Chen, and L. Chen, "Ensemble learning for healthcare analytics: A comprehensive review," Artificial Intelligence in Medicine, vol. 108, 2020.

[14] L. Huang, T. H. Chuang, and L. S. Getoor, "Stacked Survival Models for Cancer Outcome Prediction," Journal of Biomedical Informatics, vol. 112, 2020.

[15] Demographic GP. "Robust Non-Parametric Mortality and Fertility Modelling and Forecasting: Gaussian Process Regression Approaches," ResearchGate, 2021.

[16] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer, 2009.

[17] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

[18] XLSTAT, "Nonparametric Regression: Kernel and LOWESS," Addinsoft. [Online]. Available: https://www.xlstat.com/solutions/features/nonparametric-regression-kernel-and-lowess

[19] D. Ruppert and M. P. Wand, "Multivariate locally weighted least squares regression," Annals of Statistics, vol. 22, no. 3, pp. 1346–1370, 1994.

[20] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992.

[21] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[23] S. Seabold and J. Perktold, "Statsmodels: Econometric and Statistical Modeling with Python," Proceedings of the 9th Python in Science Conference, pp. 57–61, 2010.