Original Article

Check for updates

Available online at
www.heca-analitika.com/ijcr

Indonesian Journal of Case Reports

Vol. 2, No. 1, 2024

Interpretable Machine Learning for Chronic Kidney Disease Diagnosis: A Gaussian Processes Approach

Teuku Rizky Noviandy [1], Ghifari Maulana Idroes [2], Maimun Syukri [3] and Rinaldi Idroes [2,*]

[1] Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
[2] Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; ghifarimaulana145@gmail.com (G.M.I.)
[3] Department of Internal Medicine, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; maimun_62@usk.ac.id (M.S.)
[4] School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 18 April 2024
Revised 15 June 2024
Accepted 20 June 2024
Available Online 29 June 2024

Keywords:
Artificial intelligence
Medical diagnostics
Clinical decision support
SHAP

Abstract

Chronic Kidney Disease (CKD) is a global health issue impacting over 800 million people, characterized by a gradual loss of kidney function leading to severe complications. Traditional diagnostic methods, relying on laboratory tests and clinical assessments, have limitations in sensitivity and are prone to human error, particularly in the early stages of CKD. Recent advances in machine learning (ML) offer promising tools for disease diagnosis, but a lack of interpretability often hinders their adoption in clinical practice. Gaussian Processes (GP) provide a flexible ML model capable of delivering predictions and uncertainty estimates, essential for high-stakes medical applications. However, the integration of GP with interpretable methods remains underexplored. We developed an interpretable CKD classification model to address this knowledge gap by combining GP with Shapley Additive Explanations (SHAP). We assessed the model's performance using three GP kernels (Radial Basis Function, Matern, and Rational Quadratic). The results show that the Rational Quadratic kernel outperforms the other kernels, achieving an accuracy of 98.75%, precision of 100%, sensitivity of 97.87%, specificity of 100%, and an F1-score of 98.51%. SHAP values indicate that haemoglobin and specific gravity are the most influential features. The results demonstrate that the Rational Quadratic kernel enhances predictive accuracy and provides robust uncertainty estimates and interpretable explanations. This combination of accuracy and interpretability supports clinicians in making informed decisions and improving patient management and outcomes in CKD. Our study connects advanced ML techniques with practical medical applications, leading to more effective and reliable ML-driven healthcare solutions.

1. Introduction

Chronic Kidney Disease (CKD) is a global health issue that impacts millions of people worldwide, with over 800 million individuals affected, representing more than 10% of the general population [1]. CKD is characterized by a gradual loss of kidney function over time, which can lead to various complications, including cardiovascular disease, anemia, and bone disorders [2]. Early diagnosis and management of CKD are important to prevent disease progression and improve patient outcomes.

Traditionally, the diagnosis of CKD relies on laboratory tests, such as serum creatinine and estimated glomerular

filtration rate (eGFR), as well as clinical assessment by healthcare professionals [3, 4]. However, these methods have limitations, such as the lack of sensitivity in detecting early stages of CKD and the potential for human error in interpreting the results [5, 6]. Moreover, the traditional approach may not consider the complex interactions between various risk factors and biomarkers associated with CKD [7].

In recent years, machine learning (ML) has emerged as a promising tool for disease diagnosis and prediction [8–10]. ML algorithms can learn from large datasets and identify patterns that may not be apparent to human analysts [11]. ML models can provide accurate and personalized predictions for individual patients by training on diverse features, including demographic, clinical, and laboratory data. Some commonly used ML algorithms in healthcare include Support Vector Machines, Random Forests, and Artificial Neural Networks [12–14].

For instance, Qin et al. used six ML algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naïve Bayes classifier, and feedforward neural network) to predict CKD, achieving the best performance with an accuracy of 99.75% using random forest [15]. Similarly, Bai et al. employed five ML algorithms: logistic regression, naïve Bayes, random forest, decision tree, and k-nearest neighbors. The performance of each model was compared to that of the Kidney Failure Risk Equation (KFRE). Three ML models, specifically logistic regression, naïve Bayes, and random forest, demonstrated equivalent predictability and greater sensitivity than the KFRE, although the KFRE maintained the highest accuracy, specificity, and precision [16]. Rahman et al. used eight ensemble learning methods to predict CKD, employing an oversampling technique to address data imbalance and a feature selection technique. Their experimental results showed that the LightGBM method achieved the highest average accuracy (99.75%), precision (99.40%), recall (99.41%), F-measure (99.61%), and AUC-ROC (99.57%) [17]. Despite these advancements, many studies have primarily focused on model performance without providing interpretable explanations for their predictions.

One potential approach for disease diagnosis is Gaussian Processes (GP), a powerful and flexible model in ML [18]. GP is known for its ability to provide predictions and quantifiable uncertainty estimates. This characteristic is particularly valuable in the medical field, where understanding the confidence of a model's prediction can be as crucial as the prediction itself [19]. GP has demonstrated significant potential in various medical applications, including disease progression modeling and personalized treatment planning. Compared to other ML models, GPs offer the advantage of providing uncertainty measures, making them more reliable and interpretable for critical applications like healthcare.

Despite their advantages, the adoption of ML models in clinical practice is often limited by their perceived lack of interpretability [20, 21]. Clinicians require accurate predictions and insights into how these predictions are made. This necessity has fueled the rise of the interpretable ML approach, which aims to make the decision-making processes of ML models transparent and understandable to human users. One popular method for interpreting ML models is Shapley Additive Explanations (SHAP), which assigns importance values to each feature based on its contribution to the model's output [22]. SHAP has been successfully applied in various healthcare domains, including lung cancer diagnosis [23], blood glucose prediction in diabetes [24], and heart failure prediction [25].

This study aims to develop a novel framework using GP for CKD classification to enhance predictive accuracy and provide robust uncertainty estimates. It addresses the need for model interpretability by integrating SHAP, allowing clinicians to understand and trust the model's predictions. Furthermore, it bridges the gap between advanced ML techniques and practical medical applications by demonstrating the synergistic application of GP and SHAP to create interpretable and reliable diagnostic tools. This study seeks to pave the way for more effective and trustworthy ML-driven healthcare solutions by improving the performance and transparency of CKD classification models.

This study contributes to CKD diagnosis and ML in healthcare in several ways. Firstly, it introduces a novel framework that leverages GP for CKD classification, enhancing predictive accuracy and providing robust uncertainty estimates crucial for clinical decision-making. Secondly, it addresses the critical need for model interpretability by integrating SHAP, enabling clinicians to understand and trust the model's predictions. Additionally, this research bridges the gap between advanced ML techniques and practical medical applications, demonstrating how GPs and SHAP can be synergistically applied to create interpretable and reliable diagnostic tools. This study paves the way for more effective and trustworthy ML-driven healthcare solutions by improving the performance and transparency of CKD classification models.

Table 1. Features included in the dataset.

Feature	Type	Description
Age	Numerical	Age in years
Blood Pressure	Numerical	Blood pressure in mm/Hg
Specific Gravity	Categorical	Specific gravity - (1.005, 1.010, 1.015, 1.020, 1.025)
Albumin	Categorical	Albumin - (0, 1, 2, 3, 4, 5)
Sugar	Categorical	Sugar - (0, 1, 2, 3, 4, 5)
Red Blood Cells	Categorical	Red blood cells - (normal, abnormal)
Pus Cell	Categorical	Pus cell - (normal, abnormal)
Pus Cell Clumps	Categorical	Pus cell clumps - (present, not present)
Bacteria	Categorical	Bacteria - (present, not present)
Blood Glucose Random	Numerical	Blood glucose random in mg/dl
Blood Urea	Numerical	Blood urea in mg/dl
Serum Creatinine	Numerical	Serum creatinine in mg/dl
Sodium	Numerical	Sodium in mEq/L
Potassium	Numerical	Potassium in mEq/L
Haemoglobin	Numerical	Haemoglobin in gms
Packed Cell Volume	Numerical	Packed cell volume
White Blood Cell Count	Numerical	White blood cell count in cells/cumm
Red Blood Cell Count	Numerical	Red blood cell count in millions/cm
Hypertension	Categorical	Hypertension - (yes, no)
Diabetes Mellitus	Categorical	Diabetes Mellitus - (yes, no)
Coronary Artery Disease	Categorical	Coronary artery disease - (yes, no)
Appetite	Categorical	Appetite - (good, poor)
Pedal Edema	Categorical	Pedal edema - (yes, no)
Anemia	Categorical	Anemia - (yes, no)

2. Materials and Methods

2.1. Dataset

The dataset utilized in this study was sourced from the University of California Irvine (UCI) ML Repository [26]. This dataset comprises 400 instances with 24 features. Among these features, 11 are numerical, while 13 are categorical. The primary objective of this dataset is binary classification, aimed at determining whether a patient exhibits CKD or not. The specifics of the features within this dataset are presented in Table 1.

2.2. Data Preprocessing

Given that the dataset contains missing values, we use median imputation for numerical features and mode imputation for categorical features. We chose these methods because median imputation is robust to outliers, providing a central tendency measure less affected by extreme values. In contrast, mode imputation is suitable for categorical data as it preserves the most frequent category. This strategy ensures that missing values are replaced with representative statistics, maintaining the integrity of the dataset. Subsequently, we apply label encoding to prepare categorical features for ML algorithms, assigning unique numerical labels to each category [27]. This transformation facilitates the processing of categorical data, enabling algorithms to effectively interpret and utilize these features during model training.

Following imputation and encoding, we partition the dataset into training and testing sets, allocating 80% of the data for training and reserving the remaining 20% for testing. This train-test split enables us to assess the performance of our models on unseen data, providing valuable insights into their generalization capabilities. Finally, we standardize the features by removing the mean and scaling to unit variance to mitigate the impact of varying feature scales on model performance. Standardization ensures that all features have comparable scales, enhancing ML algorithms' stability and convergence. By executing these preprocessing steps, we improve the quality and suitability of the dataset for subsequent model training and evaluation [28].

2.3. Gaussian Processes

GP is a powerful and flexible ML approach for regression and classification tasks [29]. A GP is a collection of random variables, any finite number with a joint Gaussian distribution. It provides a probabilistic framework to model uncertainty in predictions, making it particularly suitable for medical applications where understanding prediction confidence is crucial [30]. GPs offer several advantages, including probabilistic predictions, kernel flexibility, and a non-parametric nature. Additionally, GPs can model various functions through different kernel functions, allowing for great flexibility in capturing data patterns. Their non-parametric nature means that GPs do not assume a fixed

form for the underlying function, making them highly adaptable to various data types.

In this study, we employ three kernel functions to assess the performance of GP for CKD classification: Radial Basis Function (RBF), Matern, and Rational Quadratic kernels. The RBF kernel, also known as the Gaussian kernel, is widely used for its simplicity and effectiveness. It defines similarity based on the distance between points in the input space, where points closer in the input space have higher similarity. The Matern kernel is a generalization of the RBF kernel and offers greater flexibility by adjusting its smoothness parameter. It is beneficial when the modeled function is expected to be less smooth. The Rational Quadratic kernel can be seen as a scale mixture of RBF kernels of different lengths and is particularly effective in scenarios where the function being modeled has varying smoothness.

2.4. Shapley Additive Explanations (SHAP)

SHAP utilizes cooperative game theory to interpret the contributions of individual features in ML models, providing a method to quantify how each feature impacts the model's prediction. In our study, we calculate SHAP values to interpret the GP model used for classifying CKD [22]. The process begins with the trained GP model, which has learned to predict CKD based on a set of features from the training dataset. We first generate predictions for the test set to calculate the SHAP values. For each prediction, we simulate the inclusion of each feature individually from a baseline that represents the absence of all features. This simulation helps understand each feature's incremental impact on the model's output [24].

We then calculate the marginal contribution of each feature by comparing the prediction with the feature to the prediction without the feature across all possible combinations of other features present in the data. This requires computing the difference in the model's output with and without each feature, averaged over all permutations of the dataset. This extensive computation accounts for interactions among features, ensuring each feature's contribution is evaluated in various contexts.

The SHAP value for each feature is the average of these differences, reflecting the feature's overall contribution to the prediction outcome [31]. Positive SHAP values indicate that the feature's presence increases the likelihood of the outcome, while negative values suggest a decrease. For instance, a positive SHAP value for haemoglobin indicates that higher haemoglobin levels are associated with an increased likelihood of CKD, providing clinical insights into the most influential features in diagnosing the disease.

2.5. Performance Evaluation

To assess the effectiveness of our model, we employ five key performance metrics: accuracy, precision, sensitivity, specificity, and F1-score [32]. Each metric provides a unique perspective on the model's capabilities and helps us understand its strengths and weaknesses in predicting CKD.

Accuracy is the proportion of correct predictions made by a model. Precision assesses the accuracy of positive predictions, indicating the share of correctly identified positive cases. Sensitivity, or the true positive rate, gauges how well the model identifies actual positive cases, which is crucial for clinical effectiveness. Specificity measures the ability to identify negatives, which is important for reducing false positives. The F1-score is a harmonic mean of precision and sensitivity, balancing correct identification of positives with avoiding false positives, useful in uneven class distributions. During our evaluation, these metrics are calculated based on the outcomes predicted by the model versus the actual labels in the test dataset, which has been set aside from the main dataset during the preprocessing phase. The equation for these four metrics is presented in Equations 1, 2, 3, 4, and 5, respectively [33].

$$Accuracy = \frac{TP + FN}{FP + FN + TP + TN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{FN + TP} \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} \quad (4)$$

$$F1 - Score = 2\frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

In these equations, TP represents the number of true positives, FN denotes the number of false negatives, FP indicates the number of false positives, and TN refers to true negatives.

Table 2. Performance metrics of different kernel functions in the GP model.

Kernel	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
RBF	97.50	100.00	95.74	100.00	97.83
Matern	97.50	100.00	95.74	100.00	97.83
Rational Quadratic	**98.75**	**100.00**	**97.87**	**100.00**	**98.51**

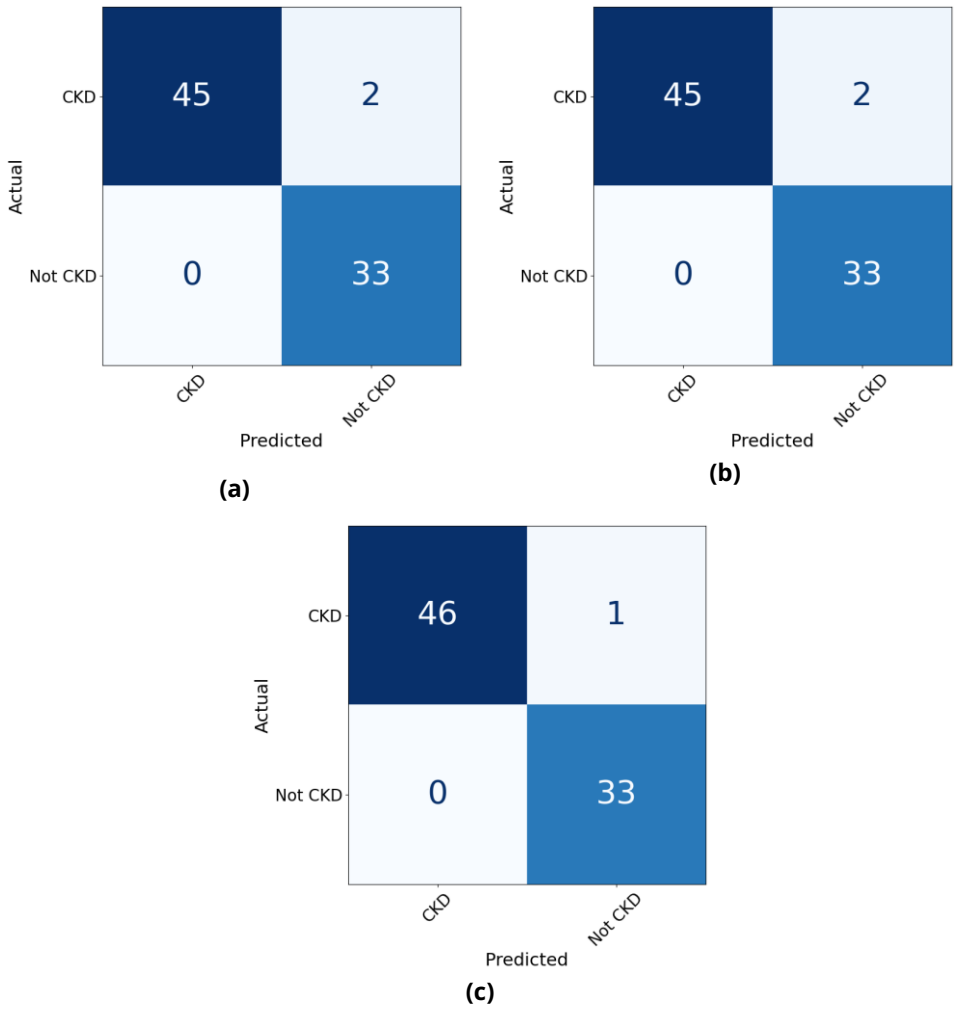

Figure 1. Confusion matrix of different kernel functions in the GP model, (a) RBF, (b) matern, and (c) rational quadratic.

3. Results and Discussion

3.1. Model Performance Evaluation

The performance of the GP model for CKD classification was evaluated using three kernel functions: RBF, Matern, and Rational Quadratic. As shown in Table 2, the Rational Quadratic kernel outperformed the other kernels, achieving an accuracy of 98.75%, higher than the 97.50% accuracy of both the RBF and Matern kernels. Precision for all kernels was 100%, indicating the model's reliability in identifying positive cases without false positives. Sensitivity was 95.74% for the RBF and Matern kernels and 97.87% for the Rational Quadratic kernel, reducing the risk of missed CKD patients. Specificity was perfect at 100% for all three kernels, ensuring accurate classification of non-CKD individuals. The F1-score was highest for the Rational Quadratic kernel at 98.51%, compared to 97.83% for the RBF and Matern kernels, indicating a better overall balance between precision and sensitivity.

The confusion matrix in Figure 1 further illustrates the performance of each kernel. The Rational Quadratic kernel correctly classified 46 out of 47 CKD cases, with only one false negative, while accurately identifying all 33 non-CKD cases. In comparison, the RBF and Matern kernels correctly classified 45 CKD cases with two false negatives, and all 33 non-CKD cases were accurately identified. These results align with the earlier performance metrics, confirming that the Rational Quadratic kernel provides the highest sensitivity and accuracy for CKD classification. The consistency in

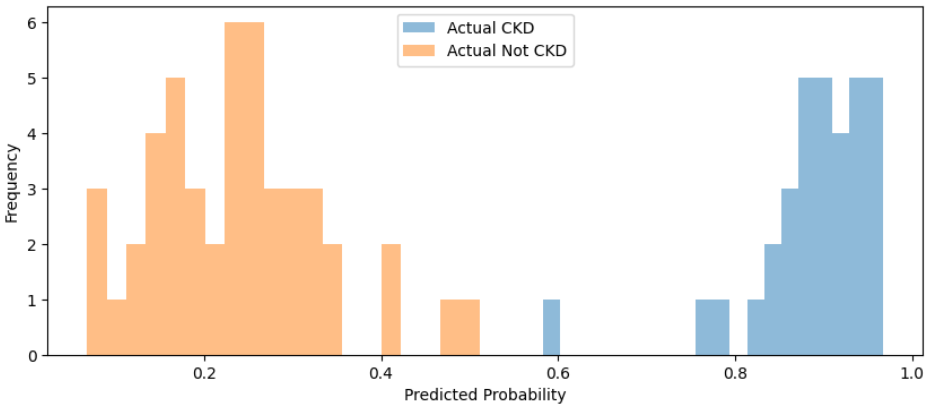

Figure 2. Histogram of predicted probabilities for CKD and non-CKD cases.

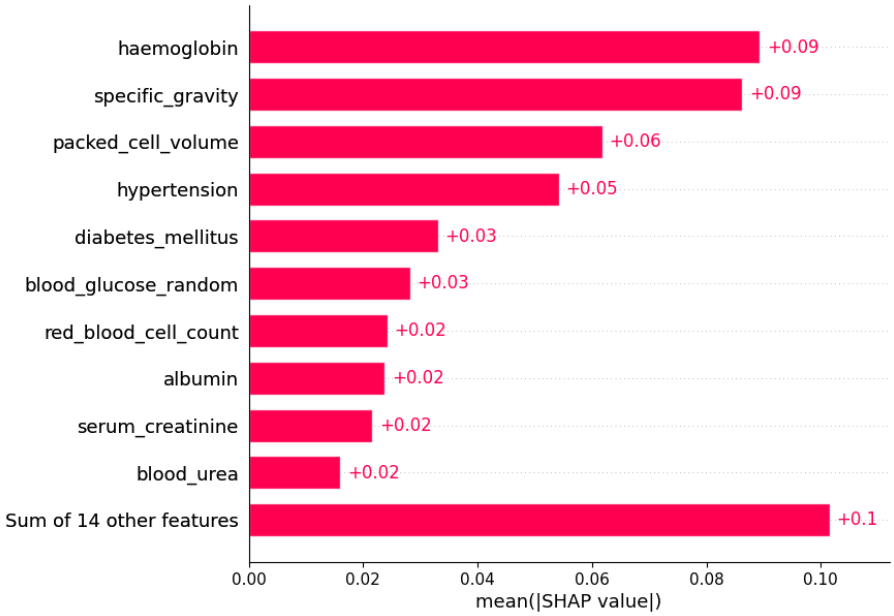

Figure 3. Mean SHAP values show feature importance in CKD classification.

specificity across all models reinforces their ability to reliably identify non-CKD individuals, ensuring robust diagnostic performance.

Figure 2 shows the distribution of predicted probabilities from a model distinguishing between CKD and non-CKD cases. The figure shows that the predicted probabilities for actual CKD cases cluster towards the higher end (around 0.8 to 1.0), indicating that the model correctly identifies most CKD cases with high confidence. Conversely, the predicted probabilities for actual non-CKD cases are mostly concentrated towards the lower end (around 0.1 to 0.4), showing that the model also successfully identifies most non-CKD cases with lower predicted probabilities.

3.2. Feature Importance and Interpretability

Figure 3 illustrates the SHAP values for the top features influencing the GP model's predictions for CKD classification. The bar plot highlights the importance of various features, with haemoglobin and specific gravity emerging as the most significant, each contributing a mean SHAP value of +0.09. Packed cell volume and hypertension follow, with SHAP values of +0.06 and +0.05, respectively. Other notable features include diabetes mellitus and blood glucose random, each contributing +0.03, while red blood cell count, albumin, serum creatinine, and blood urea each add +0.02. The cumulative impact of the remaining 14 features is represented by a mean SHAP value of +0.1. This visualization underscores the relative importance of these features in predicting CKD, providing clinicians with insights into the key factors driving the model's decisions.

Figure 4 presents a SHAP beeswarm plot, which differs from the bar plot in Figure 3 by providing a more detailed view of the distribution of SHAP values across individual predictions. Each dot in the beeswarm plot represents a single instance in the dataset, with the position on the x

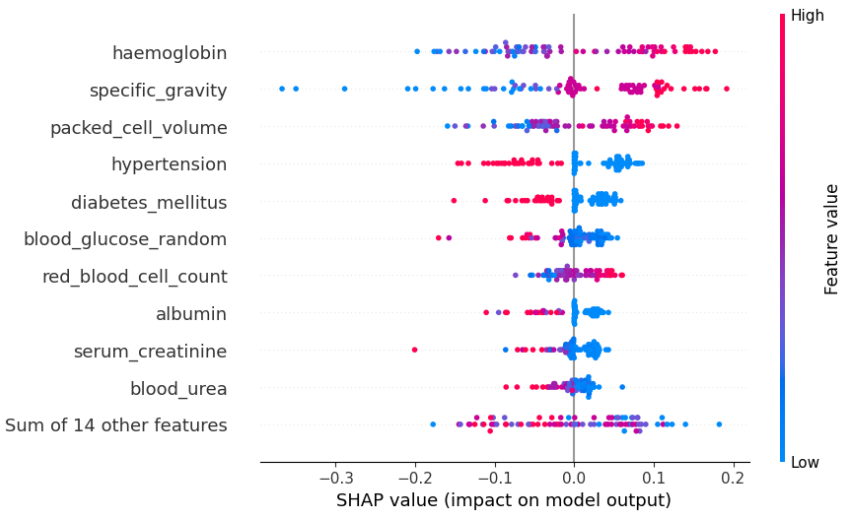

Figure 4. Beeswarm plot of SHAP values highlighting feature impact on individual predictions.

axis indicating the SHAP value's impact on the model output. The dots' colors represent the feature values, with red indicating high values and blue indicating low values.

In contrast to the summary provided by the bar plot which shows the mean SHAP values for the top features, the beeswarm plot offers a granular view, showing how the impact of each feature varies across different instances. For instance, haemoglobin and specific gravity, which had the highest mean SHAP values in Figure 3, show a wide range of SHAP values in Figure 4, indicating varying levels of impact on different predictions. This plot also reveals that features like packed cell volume and hypertension have significant impacts in many instances, with high values (in red) generally pushing the prediction toward CKD and low values (in blue) pushing it away.

Applying SHAP values provides a transparent interpretation of the model's predictions, highlighting the most influential features such as haemoglobin, specific gravity, packed cell volume, and hypertension. This interpretability is crucial in the medical field, where understanding the underlying reasons for a diagnosis can significantly enhance clinician trust and facilitate better patient management. The detailed insights from the SHAP values enable healthcare professionals to pinpoint key risk factors and biomarkers, aiding in the early detection and personalized treatment of CKD.

3.3. Practical Implication

The integration of GP with SHAP for CKD diagnosis offers several practical benefits. First, the Rational Quadratic kernel is very accurate and sensitive, which means it can, most of the time, diagnose CKD correctly. This reduces the chances of missing a diagnosis and ensures patients get the right care quickly. Second, SHAP values make the model's predictions easy to understand. Doctors can see what factors influenced the diagnosis, helping them make informed choices and create personalized treatment plans for each patient.

Clear explanations of diagnoses also improve communication between patients and doctors. When patients understand why they were diagnosed and their treatment plan, they are more likely to trust their doctors and take an active role in their care. This leads to a better relationship between patients and doctors, which is important for effective healthcare.

In addition, diagnosing CKD early and accurately can help medical resources be used more efficiently. Finding the disease early allows doctors to take steps to slow its progression and prevent complications. This proactive approach can reduce the long-term strain on the healthcare system by decreasing the need for extensive treatments and hospital stays associated with advanced CKD.

Overall, combining GP with SHAP improves diagnostic accuracy and interpretability, leading to better patient outcomes, more efficient healthcare delivery, and better resource use. This approach shows how advanced ML techniques can effectively tackle important challenges in healthcare.

3.4. Limitations and Future Directions

Despite the promising results, this study has several limitations. The dataset comprises only 400 instances, which may limit the generalizability of the findings. A larger and more diverse dataset would provide a more robust evaluation of the model's performance. While the SHAP values offer insights into feature importance, selecting features was based on existing clinical

knowledge. Future studies could explore including additional features or using automated feature selection methods to enhance model performance. Moreover, GP can be computationally intensive, particularly with larger datasets. This complexity may limit their practical application in real-time clinical settings without adequate computational resources.

Several future directions are proposed to address these limitations and further advance the field. Future research should validate the findings using larger, multi-center datasets that capture a wider range of patient demographics and clinical conditions. This would improve the model's robustness and generalizability. Developing efficient algorithms and leveraging high-performance computing resources could facilitate the real-time application of GP in clinical settings. Integrating the GP model with EHR systems could automate the diagnosis process, providing clinicians with immediate, interpretable insights during patient consultations. Conducting prospective clinical trials to evaluate the model's performance in real-world clinical settings would provide valuable evidence of its effectiveness and practical utility. Additionally, combining GP with other ML techniques, such as deep learning, could enhance predictive accuracy and model interpretability.

4. Conclusion

This study highlights the potential of using GP combined with SHAP to diagnose CKD. The Rational Quadratic kernel emerged as the most effective, delivering high accuracy, precision, sensitivity, and specificity. This kernel's superior performance underscores its ability to capture complex patterns within the CKD dataset, offering reliable diagnostic capabilities. Integrating SHAP with the GP model adds a significant layer of interpretability, which is crucial for medical applications. By providing transparent insights into feature importance, SHAP enables clinicians to understand the underlying factors driving the model's predictions. This interpretability enhances trust in the model, supports informed decision-making, and facilitates personalized patient management. Despite the promising results, the study acknowledges several limitations, including the relatively small dataset size and the computational intensity of GP. Addressing these limitations through future research involving larger, diverse datasets and efficient computational strategies will be essential. Additionally, real-world validation through clinical trials and integration of the model with electronic health record systems will further solidify its practical utility.

Author Contributions: Conceptualization, T.R.N., M.S. and R.I.; methodology, T.R.N. and R.I.; software, T.R.N. and G.M.I.; validation, M.S. and R.I.; formal analysis, T.R.N.; investigation, T.R.N. and G.M.I.; resources, R.I.; data curation, M.S. and R.I.; writing—original draft preparation, T.R.N. and G.M.I.; writing—review and editing, M.S. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study can be accessed in https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Kovesdy, C. P. (2022). Epidemiology of Chronic Kidney Disease: An Update 2022, *Kidney International Supplements*, Vol. 12, No. 1, 7–11. doi:10.1016/j.kisu.2021.11.003.
2. Levey, A. S., and Coresh, J. (2012). Chronic Kidney Disease, *The Lancet*, Vol. 379, No. 9811, 165–180. doi:10.1016/S0140-6736(11)60178-5.
3. Chen, T. K., Knicely, D. H., and Grams, M. E. (2019). Chronic Kidney Disease Diagnosis and Management, *JAMA*, Vol. 322, No. 13, 1294. doi:10.1001/jama.2019.14745.
4. Levey, A., and Inker, L. (2017). Assessment of Glomerular Filtration Rate in Health and Disease: A State of the Art Review, *Clinical Pharmacology & Therapeutics*, Vol. 102, No. 3, 405–419. doi:10.1002/cpt.729.
5. Joo, Y. S., Rim, T. H., Koh, H. B., Yi, J., Kim, H., Lee, G., Kim, Y. A., Kang, S.-W., Kim, S. S., and Park, J. T. (2023). Non-invasive Chronic Kidney Disease Risk Stratification Tool Derived from Retina-Based Deep Learning and Clinical Factors, *Npj Digital Medicine*, Vol. 6, No. 1, 114. doi:10.1038/s41746-023-00860-5.
6. Chouhan, A. S., Kaple, M., and Hingway, S. (2023). A Brief Review of Diagnostic Techniques and Clinical Management in Chronic Kidney Disease, *Cureus*. doi:10.7759/cureus.49030.
7. Arif, M. S., Mukheimer, A., and Asif, D. (2023). Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model, *Big Data and Cognitive Computing*, Vol. 7, No. 3, 144. doi:10.3390/bdcc7030144.
8. Akazawa, M., Hashimoto, K., Noda, K., and Yoshida, K. (2021). The Application of Machine Learning for Predicting Recurrence in Patients with Early-Stage Endometrial Cancer: A Pilot Study, *Obstetrics & Gynecology Science*, Vol. 64, No. 3, 266–273. doi:10.5468/ogs.20248.
9. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.
10. Solomon, D. D., Khan, S., Garg, S., Gupta, G., Almjally, A., Alabduallah, B. I., Alsagri, H. S., Ibrahim, M. M., and Abdallah, A. M. A. (2023). Hybrid Majority Voting: Prediction and Classification Model for Obesity, *Diagnostics*, Vol. 13, No. 15, 2610. doi:10.3390/diagnostics13152610.
11. Maulana, A., Faisal, F. R., Noviandy, T. R., Rizkia, T., Idroes, G. M., Tallei, T. E., El-Shazly, M., and Idroes, R. (2023). Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 1–7. doi:10.60084/ijds.v1i1.72.

12. Ali, L., Niamat, A., Khan, J. A., Golilarz, N. A., Xingzhong, X., Noor, A., Nour, R., and Bukhari, S. A. C. (2019). An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure, *IEEE Access*, Vol. 7, 54007–54014. doi:10.1109/ACCESS.2019.2909969.

13. Ahmad, N., Ul-Saufie, A. Z., Shaziayani, W. N., Abidin, A. W. Z., Zulazmi, N. E. S., and Harb, S. M. (2022). Evaluating the Performance of Random Forest and Multiple Linear Regression for Higher Observed PM10Concentrations, *Israa University Journal of Applied Science*, Vol. 6, No. 1, 72–90. doi:10.52865/WHPM9019.

14. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

15. Qin, J., Chen, L., Liu, Y., Liu, C., Feng, C., and Chen, B. (2020). A Machine Learning Methodology for Diagnosing Chronic Kidney Disease, *IEEE Access*, Vol. 8, 20991–21002. doi:10.1109/ACCESS.2019.2963053.

16. Bai, Q., Su, C., Tang, W., and Li, Y. (2022). Machine Learning to Predict End Stage Kidney Disease in Chronic Kidney Disease, *Scientific Reports*, Vol. 12, No. 1, 8377. doi:10.1038/s41598-022-12316-z.

17. Mustafizur Rahman, M., Al-Amin, M., and Hossain, J. (2024). Machine Learning Models for Chronic Kidney Disease Diagnosis and Prediction, *Biomedical Signal Processing and Control*, Vol. 87, 105368. doi:10.1016/j.bspc.2023.105368.

18. Williams, C. K. I., and Barber, D. (1998). Bayesian Classification with Gaussian Processes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, 1342–1351.

19. Liu, M., Chowdhary, G., Da Silva, B. C., Liu, S.-Y., and How, J. P. (2018). Gaussian Processes for Learning and Control: A Tutorial with Examples, *IEEE Control Systems Magazine*, Vol. 38, No. 5, 53–86.

20. Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., and Cilar, L. (2020). Interpretability of Machine Learning-Based Prediction Models in Healthcare, *WIREs Data Mining and Knowledge Discovery*, Vol. 10, No. 5. doi:10.1002/widm.1379.

21. Vellido, A. (2020). The Importance of Interpretability and Visualization in Machine Learning for Applications in Medicine and Health Care, *Neural Computing and Applications*, Vol. 32, No. 24, 18069–18083. doi:10.1007/s00521-019-04051-w.

22. Lundberg, S. M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, Vol. 30.

23. Yang, R., Xiong, X., Wang, H., and Li, W. (2022). Explainable Machine Learning Model to Prediction EGFR Mutation in Lung Cancer, *Frontiers in Oncology*, Vol. 12. doi:10.3389/fonc.2022.924144.

24. Prendin, F., Pavan, J., Cappon, G., Del Favero, S., Sparacino, G., and Facchinetti, A. (2023). The Importance of Interpreting Machine Learning Models for Blood Glucose Prediction in Diabetes: An Analysis Using SHAP, *Scientific Reports*, Vol. 13, No. 1, 16865. doi:10.1038/s41598-023-44155-x.

25. Lu, S., Chen, R., Wei, W., Belovsky, M., and Lu, X. (2021). Understanding Heart Failure Patients EHR Clinical Features via SHAP Interpretation of Tree-Based Machine Learning Model Predictions., *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, Vol. 2021, 813–822.

26. Rubini, L., Soundarapandian, P., and Eswaran, P. (2015). Chronic Kidney Disease, *UCI Machine Learning Repository*, from https://archive.ics.uci.edu/dataset/336/chronic+kidney+disease.

27. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.

28. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.

29. Rasmussen, C. E., and Nickisch, H. (2010). Gaussian Processes for Machine Learning (GPML) Toolbox, *The Journal of Machine Learning Research*, Vol. 11, 3011–3015.

30. Corbière, C., Thome, N., Bar-Hen, A., Cord, M., and Pérez, P. (2019). Addressing Failure Prediction by Learning Model Confidence, *Advances in Neural Information Processing Systems*, Vol. 32.

31. Moncada-Torres, A., van Maaren, M. C., Hendriks, M. P., Siesling, S., and Geleijnse, G. (2021). Explainable Machine Learning Can Outperform Cox Regression Predictions and Provide Insights in Breast Cancer Survival, *Scientific Reports*, Vol. 11, No. 1, 6968. doi:10.1038/s41598-021-86327-7.

32. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.

33. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Machine Learning Approach to Predict AXL Kinase Inhibitor Activity for Cancer Drug Discovery Using XGBoost and Bayesian Optimization, *Journal of Soft Computing and Data Mining*, Vol. 5, No. 1, 46–56.