Letters in Information Technology Education (LITE)
Vol 7, No 2, 2024, pp. 56-62
e-ISSN: 2654-5667

Enhancing Medical Education with Machine Learning: A Case Study on CKD Detection Using AdaBoost-SVM

Erizal Erizal[1], Suwarto Suwarto[2], Umar Basuki[3], Marselina Endah Hiswati[4], Mohammad Diqi[5,*], Tadem Vergi Kristian[6]
Universitas Respati Yogyakarta, Jl. Laksda Adisucipto Km 6.3 Sleman, Yogyakarta 55283, Indonesia
[1]erizal@respati.ac.id; [2]suwarto25nd@respati.ac.id; [3]umar.basuki@respati.ac.id; [4]marsel.endah@respati.ac.id;
[5]diqi@respati.ac.id*; [6] 20230006@respati.ac.id
* corresponding author

Article Info

Article history:
Received: Nov 5, 2024
Revised: Dec 7, 2024
Accepted: Dec 9, 2024

Keyword:
Machine Learning in Education
Chronic Kidney Disease (CKD)
AdaBoost-SVM
Medical Education
Classification Models

ABSTRACT

This study explores the application of the AdaBoost-SVM model for chronic kidney disease (CKD) detection, emphasising its relevance as a teaching tool in medical education. By leveraging a dataset of 400 instances with 25 clinical features, this research demonstrates how machine learning (ML) techniques can address the critical need for accurate and early diagnosis in clinical settings. The AdaBoost-SVM model achieved remarkable performance, with an overall accuracy of 96%. The study highlights how such models can serve as practical examples for students and educators to understand the integration of ML into healthcare better, bridging the gap between theoretical learning and real-world applications. Recommendations are provided for using this case study as an educational resource to enhance understanding of hybrid machine learning approaches in medical diagnostics.

I. INTRODUCTION

Chronic Kidney Disease (CKD) is a significant global health concern that impacts millions of people around the globe [1]. Recent figures indicate that CKD is projected to impact around 10% of the worldwide population. A considerable number of these persons are likely to advance to end-stage renal disease (ESRD), which necessitates dialysis or transplantation [2]. CKD has effects that go beyond the persons it affects, placing significant financial strain on healthcare systems and diminishing the quality of life for patients and their families [3]. As the incidence of risk factors such as diabetes, hypertension, and ageing populations increases, the prevalence of CKD is expected to rise, making it a critical area of concern for public health authorities and medical professionals alike [4].

Early detection and accurate classification of CKD are paramount in mitigating its adverse effects and improving patient outcomes [5]. Timely diagnosis allows for the implementation of therapeutic interventions that can slow disease progression, manage symptoms, and prevent complications [6]. Accurate classification of CKD stages is essential for tailoring treatment plans to individual patient needs and ensuring optimal use of healthcare resources [7]. Advances in machine learning and data analytics offer promising tools for enhancing the diagnostic accuracy and efficiency of CKD management [8]. By leveraging these technologies, healthcare providers can better identify at-risk individuals, monitor disease progression, and deliver personalised care, ultimately reducing the burden of CKD on society [9].

The classification of CKD presents significant challenges due to its complex and multifactorial nature [7]. Traditional diagnostic methods often rely on a combination of laboratory tests, clinical evaluations, and patient histories, which can be time-consuming and prone to human error [10]. Additionally, CKD symptoms can be subtle and nonspecific in the early stages, leading to delayed diagnosis and treatment [11]. Variability in clinical practices and the subjective interpretation

of test results further complicate the accurate classification of CKD [12]. These challenges underscore the urgent need for more reliable and efficient diagnostic tools that can assist healthcare professionals in identifying CKD accurately and at an early stage [13].

In this context, the integration of machine learning models into medical diagnostics holds great promise [14]. Machine learning algorithms, such as AdaBoost and Support Vector Machines (SVM), have demonstrated remarkable potential in processing large datasets, identifying patterns, and making accurate predictions [9]. These models can analyse a multitude of clinical parameters simultaneously, offering a more comprehensive and objective approach to CKD classification [15]. Effective machine learning models can enhance the precision of diagnoses, reduce the time required for assessment, and provide consistent results across diverse patient populations [16]. By leveraging these advanced computational techniques, healthcare systems can improve the early detection and management of CKD, ultimately leading to better patient outcomes and more efficient use of medical resources [17].

Current methods for classifying CKD encompass both traditional statistical techniques and innovative machine-learning approaches. Traditional methods often rely on logistic regression, decision trees, and linear discriminant analysis, which use predefined equations and criteria to categorise CKD based on clinical parameters such as glomerular filtration rate, blood pressure, and proteinuria levels [16]. While these methods have been instrumental in CKD diagnosis, they are sometimes limited by their inability to handle complex, non-linear relationships within the data [7]. On the other hand, the latest progress in machine learning provides more advanced tools, such as neural networks, random forests, and SVM. These tools are capable of handling massive datasets, discovering concealed patterns, and delivering more precise predictions [18]. These machine learning models, particularly when combined with ensemble techniques like AdaBoost, enhance diagnostic precision and allow for the integration of diverse clinical features, ultimately leading to more effective CKD management and improved patient outcomes [19].

AdaBoost, also known as Adaptive Boosting, is an effective ensemble learning method that aims to enhance the performance of weak classifiers by amalgamating them into a robust one [20]. The fundamental concept underlying AdaBoost is to train a set of weak classifiers consecutively, with each one specifically targeting the errors made by the previous classifiers. The outputs of these classifiers are then combined using a weighted majority vote [21]. This approach enhances the overall accuracy and robustness of the model [22]. In medical diagnostics, AdaBoost has been successfully applied to various tasks, such as tumour detection, disease progression prediction, and risk stratification, by effectively dealing with the complexity and variability inherent in medical data [23].

SVM is a highly prominent technique in the field of machine learning, known for its exceptional performance in solving classification issues [24]. The SVM algorithm operates by identifying the most favourable hyperplane that effectively separates data points belonging to distinct classes while maximising the distance between the hyperplane and the nearest data points [25]. This makes SVM highly effective in high-dimensional spaces and suitable for medical diagnostics, where the differentiation between classes can be subtle and complex [26]. SVMs have been widely used to diagnose diseases such as cancer, diabetes, and cardiovascular conditions, demonstrating their capability to handle large and diverse datasets with precision [27].

Combining AdaBoost and SVM leverages the strengths of both techniques, leading to enhanced classification performance [21]. Studies have shown that the AdaBoost-SVM hybrid model can significantly improve diagnostic accuracy by compensating for the individual weaknesses of each method [28]. For instance, AdaBoost's ability to focus on difficult cases and SVM's proficiency in high-dimensional classification complement each other well [27]. Research in this area has reported notable successes, such as improved detection rates in breast cancer screening and more accurate predictions of diabetic complications [20]. These findings highlight the potential of hybrid models in advancing medical diagnostics, offering more reliable and efficient tools for early disease detection and management [29].

The main aim of this study is to create a strong classification model for CKD using the AdaBoost-SVM (Support Vector Machine) method. Our objective is to utilise the benefits of AdaBoost and SVM to develop a model that can effectively and precisely detect and categorise CKD in patients using a wide range of clinical characteristics. To thoroughly assess the effectiveness of this hybrid model, we will utilize a confusion matrix and a classification report. These tools will comprehensively analyse the model's accuracy, precision, recall, and F1-score, evaluating its efficacy in differentiating between CKD and non-CKD cases.

This innovative research introduces a hybrid AdaBoost-SVM model specifically designed to classify CKD. This novel methodology combines the boosting powers of AdaBoost, which improve the performance of weak classifiers, with the robust classification capacity of SVM. Furthermore, the study highlights the importance of implementing a meticulous data-cleaning procedure to guarantee high-quality input data. Our goal is to improve the accuracy and dependability of the model by carefully managing missing values and guaranteeing proper formatting of all numeric features. This approach addresses frequent challenges in medical datasets that might impede the performance of machine learning algorithms.

Our study contributes significantly to both medical diagnostics and education by thoroughly evaluating the effectiveness of the AdaBoost-SVM model in classifying Chronic Kidney Disease (CKD) while highlighting its strengths and areas for improvement. This dual focus not only demonstrates the practical implications of hybrid machine learning models in healthcare but also emphasises their broader potential for enhancing diagnostic precision across diverse medical conditions. Positioned as both a technical achievement and a pedagogical tool, the AdaBoost-SVM methodology serves as a compelling case for interdisciplinary education, enabling educators to engage students in exploring the intersection of technology and healthcare. By analysing its application, educators can help students grasp the importance

of data-driven decision-making and develop a deeper understanding of hybrid models that combine boosting and support vector machines to address complex clinical challenges. Ultimately, this work lays a strong foundation for future innovations in machine learning applications, aiming to improve patient outcomes and advance medical education and practice.

II. METHOD

A. Dataset Description

The dataset used in this study is the Chronic Kidney Disease dataset obtained from Kaggle, which contains extensive clinical data for 400 patients. The dataset has 25 properties that include a range of medical indicators, such as red blood cell count, white blood cell count, blood pressure, specific gravity, albumin, and other pertinent laboratory test results and demographic information. These characteristics offer a comprehensive perspective on the health condition of every patient, enabling thorough examination and training of models. The goal variable, 'classification', determines whether chronic kidney disease is present ('ckd') or absent ('notckd'). It is a crucial factor in predictive modelling as it represents the predicted result. This extensive dataset enables the creation and verification of machine learning models designed to detect and categorise CKD precisely, therefore enhancing diagnostic procedures in the healthcare field.

B. Data preprocessing

Data preprocessing in this study involves meticulously handling missing values and ensuring the proper format of numerical data to enhance model accuracy. To address the issue of missing values, any row containing NaN values is entirely removed, thereby eliminating incomplete data that could skew the analysis and degrade model performance. This stringent approach ensures that the dataset remains robust and reliable. Additionally, all numeric features are converted to floating-point numbers, standardising the data format and facilitating seamless integration into the machine-learning pipeline. This step is crucial for the AdaBoost-SVM model to interpret and process the clinical parameters accurately, ultimately leading to a more precise classification of chronic kidney disease.

C. Model Development

This research uses the AdaBoost-SVM algorithm to classify CKD accurately. AdaBoost, also known as Adaptive Boosting, is a technique that combines several "weak learners" to create a powerful "strong learner." We use SVM as the weak learners in this context. The AdaBoost process involves iterative training, where each model successively focuses on the examples that the previous models misclassified.

1) AdaBoost (Adaptive Boosting)

The AdaBoost process begins with initialising the weights for each instance in the dataset. Each instance i is assigned an initial weight $w_i = \frac{1}{N}$, where N Is the total number of instances. At each iteration t, the SVM model $h_t(x)$ is trained using the current weights w_i. The weighted error of the model is calculated in Equation 1.

$$\epsilon_t = \sum_{i=1}^{N} w_i I\big(y_i \neq h_t(x_i)\big) \tag{1}$$

where I is an indicator function that is 1 if the prediction is incorrect and 0 if it is correct. Equation 2 computes the weight of the weak learner.

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) \tag{2}$$

The weights of each instance are updated based on the model's performance in Equation 3.

$$w_i \leftarrow w_i \exp\big(-\alpha_t y_i h_t(x_i)\big) \tag{3}$$

These weights are then normalized so that the total weight remains 1. The final model $H(x)$ is a weighted combination of all the weak learners, as calculated in Equation 4.

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{4}$$

2) Support Vector Machine (SVM)

SVM is a classification algorithm that seeks the optimal hyperplane, separating the data into two classes with the maximum margin. For a training dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ with $y_i \in \{-1, +1\}$, SVM minimises the objective function as calculated in Equation 5.

$$\min_{w,b} \frac{1}{2}|\boldsymbol{w}|^2 \tag{5}$$

subject to the constraint $y_i(w \cdot x_i + b) \geq 1$. For computational efficiency, this problem is often transformed into its dual form, involving Lagrange multipliers α_i as formulated in Equation 6.

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{6}$$

subject to $\sum_{i=1}^{N} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$.

3) AdaBoost-SVM

In this research, the AdaBoost-SVM algorithm combines AdaBoost's boosting process with SVM's classification strength. Each weak learner $h_t(x)$ is an SVM model trained on a subset of the data with weights provided by AdaBoost. The boosting process ensures that the SVM model in the next iteration focuses on the examples that are difficult to classify. The final model $H(x)$ Is a weighted combination of several SVM models, with each model's weight determined by its performance during training.

The AdaBoost-SVM model leverages the advantages of both algorithms to effectively manage the intricacies of medical

data, leading to enhanced precision and dependability in predicting CKD diagnosis. This combination enables the model to systematically rectify classification errors, guaranteeing that the ultimate model results from a meticulous training procedure centred on enhancing accuracy.

D. Evaluation Metrics

A confusion matrix is a tabular representation used to assess the effectiveness of a classification model by comparing the observed classifications with the expected classifications. The system comprises four essential elements: True Positive (TP) refers to the number of instances correctly identified as the positive class, such as correctly identifying patients with CKD. False Positive (FP) refers to instances incorrectly identified as the positive class, such as predicting CKD when the patient does not have it (also known as Type I error). True Negative (TN) represents the number of instances correctly identified as the negative class, such as correctly identifying patients without CKD. False Negative (FN) indicates instances incorrectly identified as the negative class, such as predicting no CKD when the patient has it (also known as Type II error).

We can calculate Precision, Recall, and F1-Score from this confusion matrix as in Equations 7-9.

$$\text{Πρεχισιον} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Ρεχαλλ} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{Φ1–Σχορε} = 2 \times \frac{\text{Πρεχισιον} \times \text{Ρεχαλλ}}{\text{Πρεχισιον} + \text{Ρεχαλλ}} \tag{9}$$

III. Result and Discussion

A. Model Performance

The effectiveness of the classification model can be comprehensively evaluated using the confusion matrix, which offers a concise and complete overview of the model's predictions compared to the actual results. Fig. 1 displays the confusion matrix, which shows the distribution of true positive, true negative, false positive, and false negative predictions for the 'ckd' and 'notckd' classes. This matrix reveals that out of 200 instances, 121 'notckd' cases were correctly identified, while only 7 needed to be misclassified. Similarly, for the 'ckd' class, the model accurately predicted 71 instances, with just 1 case incorrectly labeled. This visualisation highlights the model's accuracy and reliability in diagnosing chronic kidney disease. It underscores the importance of precise and consistent predictions in clinical decision-making, ultimately contributing to better patient outcomes and more effective healthcare interventions.

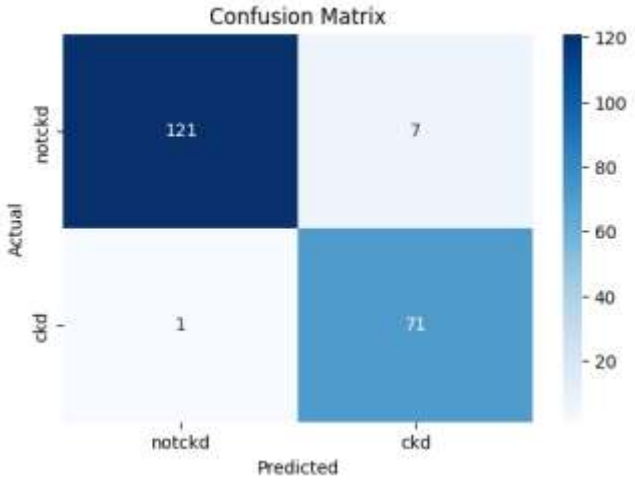

Fig. 1. Confusion matrix

The confusion matrix depicted in Fig. 1 provides a detailed evaluation of the classification model's performance in diagnosing CKD. The model demonstrated high accuracy, correctly identifying 121 out of 128 'notckd' cases and 71 out of 72 'ckd' cases. This results in a very low false positive rate of approximately 5.5% (7 out of 128) and an even lower false negative rate of about 1.4% (1 out of 72). The high true positive and true negative rates indicate that the model is highly effective in distinguishing between CKD and non-CKD patients, which is crucial for ensuring timely and appropriate medical intervention.

Moreover, these results have significant implications for clinical practice. The low false negative rate means that the likelihood of missing CKD diagnoses is minimal, reducing the risk of untreated progression of the disease. Similarly, the low false positive rate ensures that patients are not subjected to unnecessary stress or further invasive tests due to incorrect CKD diagnoses. Overall, the confusion matrix highlights the robustness of the AdaBoost-SVM model, suggesting that it can be a reliable tool for aiding healthcare professionals in the accurate and early diagnosis of CKD, ultimately leading to improved patient management and outcomes.

The detailed classification report presented in Table 1 comprehensively evaluates the model's performance across various metrics, including precision, recall, and F1-score, for both the 'notckd' and 'ckd' classes. This report highlights the model's exceptional ability to accurately predict 'notckd' cases with a precision of 0.99 and a recall of 0.95, resulting in an impressive F1-score of 0.97. Similarly, for 'ckd' cases, the model achieves a precision of 0.91 and an outstanding recall of 0.99, leading to an F1-score of 0.95. The overall accuracy of 0.96, along with macro and weighted averages of 0.96, underscores the model's robustness and reliability in diagnosing chronic kidney disease. These metrics validate the AdaBoost-SVM approach's effectiveness and demonstrate its potential to enhance clinical decision-making and patient outcomes significantly.

TABLE I. CLASSIFICATION REPORT

	Precision	Recall	F1-Score	Support
notckd	0.99	0.95	0.97	128
ckd	0.91	0.99	0.95	72
accuracy			0.96	200
macro avg	0.95	0.97	0.96	200
weighted avg	0.96	0.96	0.96	200

The classification report shown in Table 1 provides a detailed breakdown of the model's performance, illustrating its high effectiveness in classifying CKD. For the 'notckd' class, the model achieves a near-perfect precision of 0.99, indicating that almost all instances predicted as 'notckd' are correct. The recall of 0.95 for this class suggests the model successfully identifies most actual 'notckd' cases, with very few being missed. This high precision and recall combination results in an F1-score of 0.97, reflecting the model's balanced performance in minimising false positives and false negatives for the 'notckd' class.

In the case of the 'ckd' class, the model also demonstrates strong performance, with a precision of 0.91 and an outstanding recall of 0.99. The slightly lower precision indicates a small number of false positives, where non-CKD cases are incorrectly labelled as CKD. However, the extremely high recall value shows the model's exceptional ability to detect nearly all true CKD cases, ensuring that few CKD patients are overlooked. This is particularly critical in a medical context where early and accurate diagnosis is essential for effective treatment and management. The F1-score of 0.95 for the 'ckd' class signifies a well-rounded performance. Overall, the high accuracy of 0.96, along with the strong macro and weighted averages, underscores the model's reliability and robustness in real-world clinical settings, highlighting its potential to significantly improve diagnostic processes and patient outcomes in the healthcare industry.

B. Summarization of Key Findings

This research addresses the critical challenge of accurately classifying CKD using advanced machine learning techniques, specifically the hybrid AdaBoost-SVM model. The major findings indicate that the model exhibits exceptional performance, achieving an overall accuracy of 96%. For the 'notckd' class, the model attained a precision of 0.99 and a recall of 0.95, resulting in an F1-score of 0.97, while for the 'ckd' class, it achieved a precision of 0.91, a recall of 0.99, and an F1-score of 0.95. These results highlight the model's robustness and reliability in distinguishing between CKD and non-CKD cases, demonstrating its potential to enhance early diagnosis and treatment significantly, improving patient outcomes and optimising healthcare resources.

C. Result Interpretations

The results reveal high precision and recall patterns for both 'ckd' and 'notckd' classes, showcasing the AdaBoost-SVM model's ability to classify chronic kidney disease accurately. The exceptionally high recall of 0.99 for the 'ckd' class indicates that the model effectively identifies almost all true CKD cases, which is crucial for timely and accurate diagnosis. While still strong, the slightly lower precision of 0.91 for the 'ckd' class suggests a minor presence of false positives, which could be attributed to overlapping symptoms or measurement errors. These results align well with the expectations of leveraging ensemble methods to enhance classification performance. Unexpectedly, the model achieved near-perfect precision for the 'notckd' class, possibly due to the distinctiveness of non-CKD features in the dataset. Alternative explanations for these findings could involve the inherent quality and preprocessing of the dataset or the specific tuning of the model parameters, underscoring the importance of comprehensive data preparation and parameter optimisation in machine learning applications.

D. Research Implications

The implications of this research are profound, emphasising the relevance of advanced machine learning techniques in medical diagnostics and education. By successfully applying the AdaBoost-SVM model to classify chronic kidney disease (CKD) with high accuracy, this study not only confirms the effectiveness of hybrid models in improving diagnostic precision but also highlights their educational potential. The exceptional performance metrics, particularly the high recall for CKD detection, underscore the model's capability to enhance early diagnosis and patient management, aligning with the critical need for early intervention in CKD treatment. This research can serve as a foundational resource for interdisciplinary teaching, bridging the gap between technology and healthcare. By incorporating the CKD case study into educational curricula, educators can offer students hands-on experience with real-world data, enabling them to explore the practicalities of hybrid machine learning models. Such integration can enrich training programs for both technology students, who gain insights into healthcare applications, and healthcare professionals, who develop an understanding of data-driven diagnostic tools. Moreover, this approach demonstrates the importance of rigorous data cleaning and preprocessing, offering a replicable framework that students and professionals can apply to future studies or clinical practices. Ultimately, this research advances healthcare analytics, supports educational innovation, and lays the groundwork for integrating machine learning models into clinical workflows and academic settings, fostering improved patient outcomes and optimised healthcare resources.

E. Research Limitations

Although this study offers new insights into the categorisation of chronic renal illness using the AdaBoost-SVM model, it is crucial to recognise its limitations. The primary constraint is in the minimal and particular dataset, which may not encompass the complete range of variations observed in more significant and more diverse populations. This could impact the generalizability of the results. Additionally, the strict data cleaning process, which involved removing all rows with missing values, may have excluded potentially informative cases. Despite these limitations, the high accuracy and strong performance metrics indicate that the model effectively addresses the research question, demonstrating its potential for accurate CKD classification. The robust methodology and consistent results suggest that the findings are valid and can inform future research and applications while highlighting areas

where further investigation and broader data collection could enhance the model's applicability and reliability.

F. Recommendations for Future Research

To effectively deploy the AdaBoost-SVM model, it is advisable to incorporate it into clinical decision support systems. This integration will assist healthcare workers in diagnosing chronic kidney disease promptly and accurately. Future research should focus on utilising larger, more diverse datasets encompassing a wider range of demographic and clinical variations to enhance its applicability. Additionally, exploring different data imputation techniques to handle missing values rather than simply removing incomplete rows could preserve valuable information and improve model performance. Further studies could also investigate integrating other advanced machine learning algorithms and ensemble methods to refine the model's accuracy and robustness. Ultimately, collaboration between data scientists and healthcare practitioners will be crucial in tailoring these models to real-world clinical settings, ensuring they provide actionable insights and improve patient outcomes.

IV. CONCLUSION

In conclusion, this study demonstrates the effectiveness of the AdaBoost-SVM model in accurately classifying chronic kidney disease, achieving high precision, recall, and overall accuracy while also showcasing its potential as an educational tool. The rigorous data cleaning and preprocessing steps ensured the reliability of the input data, contributing to the model's robust performance and making it an exemplary case for interdisciplinary teaching. Despite some limitations, such as the use of a relatively small dataset and the exclusion of incomplete data, the findings offer valuable insights into the application of hybrid machine learning models in enhancing both medical diagnostics and educational practices. This research not only validates the application of advanced algorithms in healthcare but also highlights their relevance in academic settings, encouraging the integration of such models into curricula for medical and technology students. By paving the way for future studies and educational innovations, this work aims to improve early diagnosis, patient management, and the training of future professionals in data-driven healthcare solutions.

References

[1] P. Cockwell and L. Fisher, "The Global Burden of Chronic Kidney Disease," *The Lancet*, vol. 395, no. 10225, pp. 662–664, 2020, doi: 10.1016/s0140-6736(19)32977-0.

[2] J. G. R. de Oliveira, M. Askari, G. B. da Silva, R. A. de F. Filho, and J. E. V. Filho, "Renal Health: An Innovative Application to Increase Adherence to Treatment Through Self-Monitoring for Patients With CKD and Provide Information for the General Population," *Kidney Int. Rep.*, vol. 4, no. 4, pp. 609–613, 2019, doi: 10.1016/j.ekir.2019.01.008.

[3] B. Bikbov *et al.*, "Global, Regional, and National Burden of Chronic Kidney Disease, 1990–2017: A Systematic Analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 395, no. 10225, pp. 709–733, 2020, doi: 10.1016/s0140-6736(20)30045-3.

[4] E. F. Carney, "The Impact of Chronic Kidney Disease on Global Health," *Nat. Rev. Nephrol.*, vol. 16, no. 5, pp. 251–251, 2020, doi: 10.1038/s41581-020-0268-7.

[5] M. J. Valente *et al.*, "P0739panel of Sensitive Biomarkers of the Primary Response to Renal Injury for an Early Diagnosis of Chronic Kidney Disease," *Nephrol. Dial. Transplant.*, vol. 35, no. Supplement_3, 2020, doi: 10.1093/ndt/gfaa142.p0739.

[6] M. G. Shlipak *et al.*, "The Case for Early Identification and Intervention of Chronic Kidney Disease: Conclusions From a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference," *Kidney Int.*, vol. 99, no. 1, pp. 34–47, 2021, doi: 10.1016/j.kint.2020.10.012.

[7] S. Eddy, L. Mariani, and M. Kretzler, "Integrated Multi-Omics Approaches to Improve Classification of Chronic Kidney Disease," *Nat. Rev. Nephrol.*, vol. 16, no. 11, pp. 657–668, 2020, doi: 10.1038/s41581-020-0286-5.

[8] L. Antony *et al.*, "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction," *Ieee Access*, vol. 9, pp. 126481–126501, 2021, doi: 10.1109/access.2021.3109168.

[9] G. Chen *et al.*, "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform," *Ieee Access*, vol. 8, pp. 100497–100508, 2020, doi: 10.1109/access.2020.2995310.

[10] F. Trevisani *et al.*, "Renal Histology Across the Stages of Chronic Kidney Disease," *J. Nephrol.*, vol. 34, no. 3, pp. 699–707, 2021, doi: 10.1007/s40620-020-00905-y.

[11] L. J. F. Peters, J. Floege, E. A. L. Biessen, J. Jankowski, and E. P. C. van der Vorst, "MicroRNAs in Chronic Kidney Disease: Four Candidates for Clinical Application," *Int. J. Mol. Sci.*, vol. 21, no. 18, p. 6547, 2020, doi: 10.3390/ijms21186547.

[12] C. Weber *et al.*, "Optimized Identification of Advanced Chronic Kidney Disease and Absence of Kidney Disease by Combining Different Electronic Health Data Resources and by Applying Machine Learning Strategies," *J. Clin. Med.*, vol. 9, no. 9, p. 2955, 2020, doi: 10.3390/jcm9092955.

[13] M. Elhoseny, K. Shankar, and J. Uthayakumar, "Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease," *Sci. Rep.*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-46074-2.

[14] S. Nusinovici *et al.*, "Logistic Regression Was as Good as Machine Learning for Predicting Major Chronic Diseases," *J. Clin. Epidemiol.*, vol. 122, pp. 56–69, 2020, doi: 10.1016/j.jclinepi.2020.03.002.

[15] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A Machine Learning Methodology for Diagnosing Chronic Kidney Disease," *Ieee Access*, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/access.2019.2963053.

[16] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *Ieee Access*, vol. 9, pp. 17312–17334, 2021, doi: 10.1109/access.2021.3053763.

[17] H. Δρίτσας and M. Trigka, "Machine Learning Techniques for Chronic Kidney Disease Risk Prediction," *Big Data Cogn. Comput.*, vol. 6, no. 3, p. 98, 2022, doi: 10.3390/bdcc6030098.

[18] N. AlMansour *et al.*, "Neural Network and Support Vector Machine for the Prediction of Chronic Kidney Disease: A Comparative Study," *Comput. Biol. Med.*, vol. 109, pp. 101–111, 2019, doi: 10.1016/j.compbiomed.2019.04.017.

[19] M. A. Abdel-Fattah, N. A. Othman, and N. Goher, "Predicting Chronic Kidney Disease Using Hybrid Machine Learning Based on Apache Spark," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/9898831.

[20] Q. Huang, Y. Chen, L. Liu, D. Tao, and X. Li, "On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification," *Ieee Trans. Knowl. Data Eng.*, vol. 32, no. 4, pp. 728–738, 2020, doi: 10.1109/tkde.2019.2891622.

[21] J. Zheng, D. Lin, Z. Gao, S. Wang, M. He, and J. Fan, "Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis," *Ieee Access*, vol. 8, pp. 96946–96954, 2020, doi: 10.1109/access.2020.2993536.

[22] S.-J. Tu, V. T. Tran, J. M. Teo, W. C. Chong, and J.-R. Tseng, "Utility of Radiomic Zones for Risk Classification and Clinical Outcome Predictions Using Supervised Machine Learning During Simultaneous 11C-choline PET/MRI Acquisition in Prostate Cancer Patients," *Med. Phys.*, vol. 48, no. 9, pp. 5192–5201, 2021, doi: 10.1002/mp.15064.

[23] N. S. M. Lak *et al.*, “Improving Risk Stratification for Pediatric Patients With Rhabdomyosarcoma by Molecular Detection of Disseminated Disease,” *Clin. Cancer Res.*, vol. 27, no. 20, pp. 5576–5585, 2021, doi: 10.1158/1078-0432.ccr-21-1083.

[24] D. Pisner and D. M. Schnyer, “Support Vector Machine,” pp. 101–121, 2020, doi: 10.1016/b978-0-12-815739-8.00006-7.

[25] Y. Jiang *et al.*, “Tumor Immune Microenvironment and Chemosensitivity Signature for Predicting Response to Chemotherapy in Gastric Cancer,” *Cancer Immunol. Res.*, vol. 7, no. 12, pp. 2065–2073, 2019, doi: 10.1158/2326-6066.cir-19-0311.

[26] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, “Discovering the Type 2 Diabetes in Electronic Health Records Using the Sparse Balanced Support Vector Machine,” *Ieee J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 235–246, 2020, doi: 10.1109/jbhi.2019.2899218.

[27] M. Wang and H. Chen, “Chaotic Multi-Swarm Whale Optimizer Boosted Support Vector Machine for Medical Diagnosis,” *Appl. Soft Comput.*, vol. 88, p. 105946, 2020, doi: 10.1016/j.asoc.2019.105946.

[28] S. K. Kalagotla, S. V. Gangashetty, and K. Giridhar, “A Novel Stacking Technique for Prediction of Diabetes,” *Comput. Biol. Med.*, vol. 135, p. 104554, 2021, doi: 10.1016/j.compbiomed.2021.104554.

[29] S. Roy, S. Das, D. Kar, F. Schwenker, and R. Sarkar, “Computer Aided Breast Cancer Detection Using Ensembling of Texture and Statistical Image Features,” *Sensors*, vol. 21, no. 11, p. 3628, 2021, doi: 10.3390/s21113628.