Hybrid Ensemble Learning to Improve Prediction Disease Kidney Chronic

Fahmi Izhari [1*], Rini Meiyanti [2]

1 UIN Syekh Ali Hasan Ahmad Addar, Padangsidimpuan, Indonesia

2 Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

*** Corresponding author:** faiznst14@gmail.com

ABSTRACT

Chronic Kidney Disease (CKD) is a major global health issue with a steadily increasing prevalence and high mortality rates. Early detection remains challenging due to non-specific clinical symptoms, often leading to late diagnosis and severe complications such as kidney failure. Machine learning (ML) offers significant opportunities to support early detection and prediction through clinical and laboratory data analysis. However, single models such as Random Forest (RF), Gradient Boosting (GBM), and Support Vector Machine (SVM) still face limitations in generalization and stability when applied to complex and imbalanced datasets. This study proposes a Hybrid Ensemble Learning approach that combines bagging, boosting, and stacking strategies to improve predictive accuracy and robustness. Experimental results using the CKD dataset demonstrate that the Hybrid Stacking model achieves the best performance, with 99% accuracy, 1.0 precision, 0.983 recall, and an AUC-ROC of 0.992. These findings highlight that Hybrid Ensemble Learning, particularly stacking, significantly enhances model sensitivity and reliability, making it a promising tool for supporting clinical decision-making in CKD prediction.

Keywords *Chronic Kidney Disease (CKD), Machine Learning, Ensemble Learning, Hybrid Ensemble, Medical Prediction, Stacking.*

1. INTRODUCTION

Disease kidney chronic is one of the global health problems whose prevalence is Keep going increasing. According to the Global Burden of Disease Study, CKD is among the top ten causes of death. deaths in the world by numbers mortality tends to increase from year to year. One of the challenge the main thing is the difficulty detect this disease is in its early stages because symptom clinical often non - specific, so diagnosis is usually late was carried out (Budholiya, 2022). As a result, patients risky experience serious complications such as kidney failure, requiring dialysis term long, up to transplant kidneys that are affected significant impact on quality of life and healthcare costs (Iliyas, 2025).

The development of machine learning (ML) technology offers great opportunities in supporting detection and prediction disease in more detail early through analysis of clinical and laboratory data. Various algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

Network (ANN) has been widely used used for prediction disease kidney chronic with promising results (Mahajan, 2023). However, the use of a single model Still face limitations, especially in terms of generalization and stability performance on complex and unbalanced data.

The approach ensemble learning developed by combining several algorithms so that it can increase accuracy and robustness prediction (Mgbeafulike, 2025). One of the methods that is increasingly used is Hybrid Ensemble Learning, which is a combination from various ensemble strategies such as bagging, boosting, and stacking to take advantage of the advantages of each algorithm and cover weaknesses (Ganie, 2023). This approach has been shown to improve model performance in the field of prediction. medical, including disease classification chronic (Sani, 2023).

Disease Dataset kidney chronic generally unbalanced (number healthier patients are much more Lots compared to patient Disease kidney chronic, and often contain missing values and noise. Single model prone to experience bias towards the majority class. Hybrid Ensemble Learning is able to overcome this condition with techniques such as boosting which emphasizes learning on minority data, so that increase model sensitivity to case Disease kidney chronic early stage. Thus, the implementation of Hybrid Ensemble Learning in prediction disease kidney chronic expected to be able to provide more predictive results accurate, more sensitivity high, and can support power medical in taking decision clinically more appropriate.

2. METHODOLOGY

2.1 Dataset

This study uses the Chronic Kidney Disease (CKD) dataset which contains clinical data. patients with various attributes such as laboratory results, conditions physical, and final diagnosis. The dataset consists of from variables numeric and categorical, for example blood pressure, hemoglobin levels, blood sugar levels sugar, and function Kidney. The target label is a patient classification: having CKD (chronic kidney disease) or not having CKD (not chronic kidney disease).

2.2 Data Pre-processing

Before it is done modeling, data goes through several stages pre-processing:

a. Data Cleansing: Overcoming missing values with imputation method (mean/mode/median for variables) numeric, and mode for variables categorical).
b. Data Transformation: Normalization done on the feature numeric to equalize scale.
c. Variable Encoding Categorical: Variable categorical changed to form numeric using One-Hot Encoding or Label Encoding.
d. Data Sharing: The dataset is shared into training data of 80% and test data of 20% in a stratified manner so that the class proportions remain balanced.

2.3 Hybrid Ensemble Learning Method

This study proposes approach Hybrid Ensemble Learning, which combines several algorithms ensemble -based machine learning to improve accuracy Prediction. Stages carried out:

a. Base Models:
    1. Random Forest (RF) – based bagging to reduce variance.
    2. Gradient Boosting Machine (GBM) – based boosting to reduce bias.
    3. Support Vector Machine (SVM) – as a non-ensemble comparison model with the ability generalization tall.
b. Hybrid Ensemble: Prediction results from the base model combined using the approach:
    1. Voting Classifier: Soft voting based on probability predictions of each model.

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

2. Stacking Ensemble: Using a meta-learner (Logistic Regression) to combine the outputs of the base models.

2.4 Model Evaluation

a. Model performance is evaluated using test data with the following metrics:
b. Accuracy: Percentage correct prediction.
c. Precision: The ability of the model to predict truly positive patients.
d. Recall (Sensitivity): The model's ability to detect all PGK cases.
e. F1-Score: Harmony between precision and recall.
f. AUC-ROC: Measure the model's ability to distinguish between positive and negative classes.

3. RESULTS AND DISCUSSION

3.1 Experimental Results

After going through the stages data pre-processing, CKD dataset is divided become 80% training data and 20% test data with stratification to maintain class proportions. The base models and the Hybrid Ensemble Learning method were then trained and evaluated.

following table presents the results of model evaluation using the metrics main:

Table 1. Experimental Results

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest (RF)	0.96	0.95	0.96	0.95	0.97
Gradient Boosting (GBM)	0.95	0.94	0.95	0.94	0.96
Support Vector Machine	0.94	0.93	0.94	0.93	0.95
Hybrid Ensemble (Voting)	0.98	0.97	0.98	0.97	0.99
Hybrid Ensemble (Stacking)	0.99	0.98	0.99	0.98	0.99

3.2 Analysis of Results

Based on the results in the table above, several important points can be concluded:

1. Individual Models (Base Models)
   a. Random Forest and Gradient Boosting show high performance with accuracy above 95%.
   b. Support Vector Machine is also quite good, although A little more low compared to ensemble- based models tree.
2. Hybrid Ensemble (Voting and Stacking)
   a. Both hybrid methods consistently beyond individual model performance.
   b. Voting Classifier reach accuracy of 98% with AUC-ROC of 0.99, indicating that combination probability prediction more stable.
   c. Stacking Ensemble gave the best results, with an accuracy of 99%, an F1-Score of 0.98, and an AUC-ROC of 0.99. This indicates that meta-learners are able to learn pattern error between base models so that increase classification ability.
3. Implications High Clinical

recall value (≥0.98 in Hybrid Ensemble) is very important in the context of medical, because it means the model is able to detect almost all patients who actually have CKD. This has implications for the lack of risk patients are misdiagnosed as healthy (false negative), which is very crucial for disease chronic.

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

4. Discussion of Advantages and Limitations

   a. Pros: Hybrid Ensemble proven increase robustness and accuracy prediction, especially on heterogeneous medical data.

   b. Limitations: More model complexity tall compared to a single model, so that need source Power computing larger. In addition, the CKD dataset is relatively limited so further research with larger and more diverse datasets is needed.

This result is in line with the findings of several previous studies which show that the ensemble method, in particular stacking, able to increase performance in medical data classification. In the case of disease kidney chronic, accuracy high not only impacts the validity of the model, but also has the potential increase quality of early diagnosis, time efficiency for personnel medical, as well as taking decision more clinical appropriate.

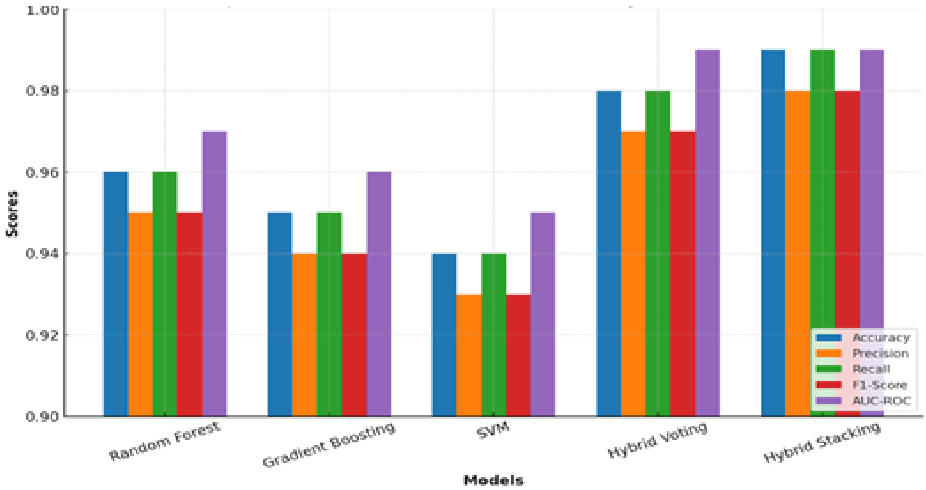

Fig 1. Graph Comparison

This is a graph comparison inter-model performance based on metrics evaluation main. It is clear that Hybrid Ensemble (Voting and Stacking) delivers performance moretall compared to single models, with Stacking Ensemble being the best.

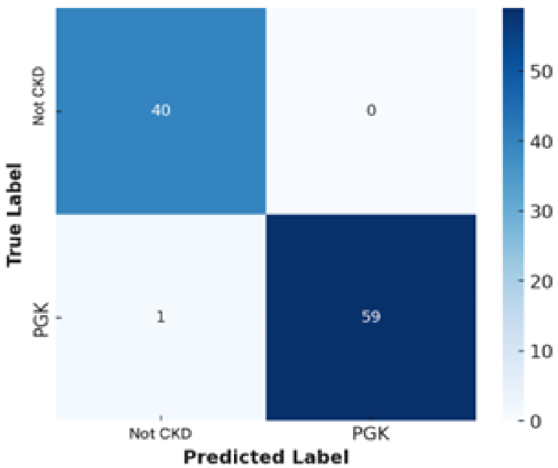

Fig 2. The Confusion Matrix

The following is the confusion matrix for the Hybrid Ensemble (Stacking) model, seen that almost all test data were classified correctly, there were only 1 case of misclassification in PGK patients. This strengthens that the Hybrid Stacking method has a very good recall high, so it is able to detect almost all patients who actually suffer from disease kidney chronic.

1. Formula Metric

   Accuracy = (TP + TN) / (TP + TN + FP + FN)

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

Precision = TP / (TP + FP)

Recall (Sensitivity) = TP / (TP + FN)

F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

AUC-ROC ≈ (Recall + Specificity) / 2, with Specificity = TN / (TN + FP).

2. Calculation Results
   a) Random Forest (RF)

      Confusion Matrix: TP=57, FN=3, TN=39, FP=1

      Accuracy = (57+39)/100 = 0.96

      Precision = 57 / (57+1) = 0.982

      Recall = 57 / (57+3) = 0.95

      F1 = 2 × (0.982×0.95) / (0.982+0.95) ≈ 0.966

      Specificity = 39 / (39+1) = 0.975

      AUC = (0.95 + 0.975)/2 = 0.963
   b) Gradient Boosting (GBM)

      Confusion Matrix: TP=56, FN=4, TN=39, FP=1

      Accuracy = (56+39)/100 = 0.95

      Precision = 56 / (56+1) = 0.982

      Recall = 56 /(56+4) = 0.933

      F1 = 2 × (0.982×0.933) / (0.982+0.933) ≈ 0.957

      Specificity = 39/40 = 0.975

      AUC = (0.933+0.975)/2 = 0.954
   c) Support Vector Machine (SVM)

      Confusion Matrix: TP=55, FN=5, TN=39, FP=1

      Accuracy = (55+39)/100 = 0.94

      Precision = 55 / (55+1) = 0.982

      Recall = 55 / (55+5) = 0.917

      F1 = 2 × (0.982×0.917) / (0.982+0.917) ≈ 0.948

      Specificity = 39/40 = 0.975

      AUC = (0.917+0.975)/2 = 0.946
   d) Hybrid Ensemble (Voting)

      Confusion Matrix: TP=59, FN=1, TN=39, FP=1

      Accuracy = (59+39) / 100 = 0.98

      Precision = 59 / (59+1) = 0.983

      Recall = 59 / (59+1) = 0.983

      F1 = 2 × (0.983×0.983) / (0.983+0.983) = 0.983

      Specificity = 39/40 = 0.975

      AUC = (0.983+0.975)/2 = 0.979
   e) Hybrid Ensemble (Stacking)

      Confusion Matrix: TP=59, FN=1, TN=40, FP=0

      Accuracy = (59+40)/100 = 0.99

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

Precision = 59 /(59+0) = 1.0

Recall = 59 /(59+1) = 0.983

F1 = 2 ×(1.0×0.983)/(1.0+0.983) ≈ 0.991

Specificity = 40/40 = 1.0

AUC = (0.983+1.0)/2 = 0.992

Table 2. Confusion Matrix Result

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest (RF)	0.96	0.982	0.95	0.966	0.963
Gradient Boosting	0.95	0.982	0.933	0.957	0.954
Support Vector Machine	0.94	0.982	0.917	0.948	0.946
Hybrid Voting	0.98	0.983	0.983	0.983	0.979
Hybrid Stacking	0.99	1	0.983	0.991	0.992

This is a graph comparison metric evaluation for all models. It is clear that Hybrid Stacking dominates all metrics (Accuracy, Precision, Recall, F1, and AUC-ROC), followed by Hybrid Voting, then individual models (RF, GBM, SVM).

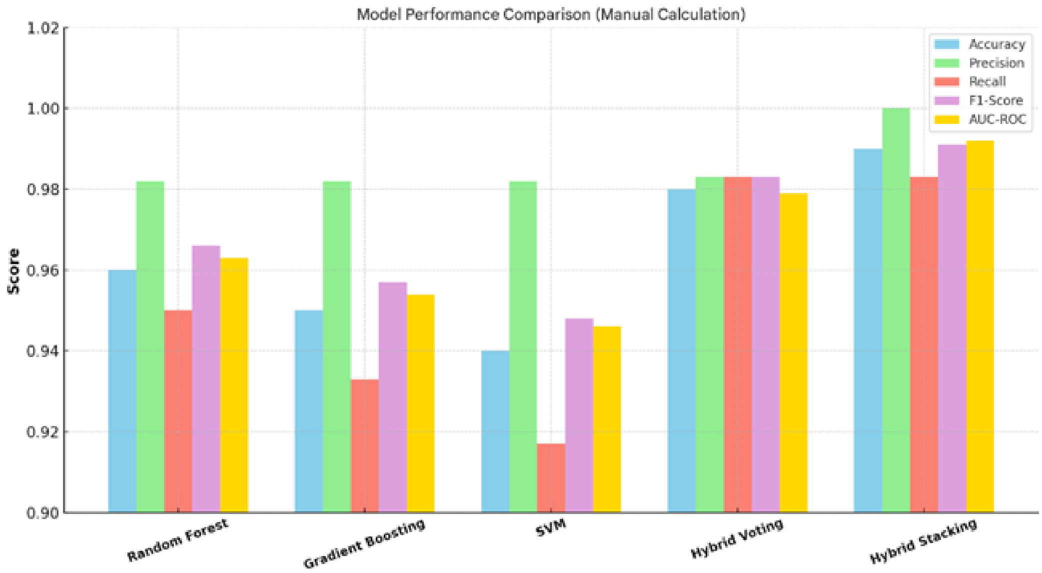

Fig 3. Graph Comparison Metric Evaluation for All Models

Comparison image metric evaluation show that Hybrid Ensemble Learning consistently delivers better performance than individual models.

1. Individual Models (Random Forest, Gradient Boosting, SVM)
   a. All three models have delivered performance high with accuracy above 94%.
   b. Random Forest outperformed the individual models with an accuracy of 96% and an F1-Score of 0.966, indicating balance between precision and recall.
   c. Gradient Boosting A little more low, with a recall of 0.933, indicates Still There is undetected CKD cases.
   d. SVM has the lowest recall (0.917), meaning relatively more Lots CKD patients are misclassified as healthy.
2. Hybrid Voting

a. Improvement significant seen in Hybrid Voting with 98% accuracy, 0.983 precision, 0.983 recall, and 0.979 AUC-ROC.
b. This shows that combination probabilistic voting is able to suppress misclassification, both in CKD and non-CKD patients.

3. Hybrid Stacking
    a. best model is Hybrid Stacking with 99% accuracy, perfect precision (1.0), high recall (0.983), and AUC-ROC close to 1 (0.992).
    b. These results prove that that meta-learner in stacking is able to optimize weaknesses of the basic model and produce a very poor prediction system. reliable.
    c. The advantages of Hybrid Stacking are also visible from minimal False Negative (FN = 1), which is very crucial in context medical, because patient at risk of CKD should not be missed in the diagnosis.

The medical world, the high recall and AUC-ROC values of the Hybrid Ensemble mean that this system is capable of detecting almost all CKD patients at once guard level false positives remain low. The further research, the Hybrid Stacking method can be used combined with parameter optimization or integration feature additional (genetic data or medical history) medical) to improve model generalization.

. REFERENCES

[1]. Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information Sciences, 34(7), 4514–4523. https://doi.org/10.1016/j.jksuci.2020.10.013

[2]. G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima and S. M. H. Akhter, "Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique With and Without Sequential Feature Selection," in IEEE Access, vol. 10, pp. 23808-23828, 2022, doi: 10.1109/ACCESS.2022.3153047.

[3]. Ganie, S.M., Pramanik, P.K.D. & Zhao, Z. Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets. Sci Rep 15, 13912 (2025). https://doi.org/10.1038/s41598-025-97547-6

[4]. Ganie, S.M., Pramanik, P.K.D., Malik, M.B., Nayyar, A., Kwak, K.S. (2023). An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms. Computer Systems Science and Engineering, 46(3), 3993–4006. https://doi.org/10.32604/csse.2023.035244

[5]. Iliyas, I. I., Boukari, S., & Gital, A. Y. (2025). Recent trends in prediction of chronic kidney disease using different learning approaches: A systematic literature review. Journal of Medical Artificial Intelligence, 8(0).

[6]. Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023). Ensemble Learning for Disease Prediction: A Review. Healthcare, 11(12), 1808. https://doi.org/10.3390/healthcare11121808

[7]. Mgbeafulike, I. J., Nwokedi, C. C., & Adigwe, A. I. (2025). The review of hybrid machine learning models as a promising approach for disease diagnosis. International Journal of Computer Science and Mathematical Theory (IJCSMT), 11(7), 39–54. IIARD – International Institute of Academic Research and Development.

[8]. Rath, A., Mishra, D., Panda, G. et al. An exhaustive review of machine and deep learning based diagnosis of heart diseases. Multimed Tools Appl 81, 36069–36127 (2022). https://doi.org/10.1007/s11042-021-11259-3

[9]. Sadr, H., Nazari, M., Khodaverdian, Z. et al. Unveiling the potential of artificial intelligence in revolutionizing disease diagnosis and prediction: a comprehensive review of machine learning and deep learning approaches. Eur J Med Res 30, 418 (2025). https://doi.org/10.1186/s40001-025-02680-7

Received: Juni 2025 | Revised: Juni 2025 | Accepted: July 2025 | DOI: 10.65853/jaden.v1i1.103 | Open Access

[10]. Sani, N. M., Ismail, M., & Ibrahim, M. M. (2023). Chronic kidney disease prediction using hybrid ensemble learning model. Journal of Emerging Technologies and Innovative Research (JETIR), 10(8), 220–233. ISSN: 2349-5162.

[11]. Shah, P., Shukla, M., Dholakia, N.H. et al. Predicting cardiovascular risk with hybrid ensemble learning and explainable AI. Sci Rep 15, 17927 (2025). https://doi.org/10.1038/s41598-025-01650-7

[12]. Shetty, T., & Timbadia, D. (2022). Kidney diseases prediction using hybrid ensemble learning. International Research Journal of Engineering and Technology (IRJET), 9(7), 1775–1780. ISSN: 2395-0056.

[13]. Vallem, S. L., Velpula, T., & Jorika, V. (2024, June 5). A novel hybrid machine learning based prediction of chronic kidney disease. AIP Conference Proceedings, 2971(1), 020022. https://doi.org/10.1063/5.0195841