Diagnosis and Prediction of Chronic Kidney Disease Using a Stacked Generalization Approach

Agung Prabowo[1] Sumita Wardani[2] Abdul Muis [3] Radiman Gea[4] Nathanael Atan Baskita Tarigan [5]
[1]Universitas Prima Indonesia [2]Universitas Satya Terra Bhinneka
[1]agungprabowo@unprimdn.ac.id, [2]sumitawardani@unprimdn.ac.id, [3]abdulmuis@satyaterrabhinneka.ac.id
[4]radimangea237@gmail.com, [5]natanaelatan5@gmail.com

ABSTRACT

Chronic Kidney Disease (CKD) is. In the past, several learners have been applied for prediction of CKD but there is still enough space to develop classifiers with higher accuracy. The study utilizes chronic kidney disease dataset from UCI Machine Learning Repository. In this paper, individual approaches, viz., linear-SVM, kernel methods including polynomial, radial basis function, and sigmoid have been used while among ensembles majority voting and stacking strategies have been applied. Stacked Ensemble is based on various types of meta-learners such as C4.5, NB, k-NN, SMO, and logit-boost. The stacking approach with meta-learner Logit-Boost (ST-LB) achieves accuracy 98,50%, sensitivity 98,50%, false positive rate 20,00%, precision 98,50%, and F-measure 98,50% demonstrating that it is the best classifier as compared to any of the individual and ensemble approaches

Keywords: Chronic Kidney Disease, Classification, Machine Learning, Majority Voting, Ensemble, Stacked Generalization, Support Vector Machine

INTRODUCTION

One of the global health issues that poses a significant burden on individuals, healthcare systems, and economies is chronic kidney disease (Hill, 2022). Chronic kidney disease can occur when the kidneys undergo gradual damage and permanent loss of function, thus leading to the need for kidney replacement therapies such as dialysis or transplantation (Bikbov, 2020). Currently, chronic kidney disease can be prevented or treated effectively if detected early. The greatest challenge in clinical practice is how this disease can be detected and diagnosed early (Smith, 2023).

The current diagnostic methods for chronic kidney disease primarily rely on physical examination, blood tests, and urine tests, which may not be sufficiently sensitive or specific to detect chronic kidney disease at an early stage (Gupta, 2023). On the other hand, the prediction of chronic kidney disease progression often relies on patient risk factors and clinical parameters, which are often difficult to interpret accurately without adequate technological support (Wang, 2022)

In recent years, advancements in the field of machine learning, particularly the Stacked Generalization approach, have shown great potential in enhancing the accuracy and robustness of predictive models (Geron, 2019). This approach involves combining predictions from various machine learning models to generate more accurate predictions. In the context of chronic kidney disease, the Stacked Generalization approach can be utilized to integrate information from various clinical and laboratory features to enhance the prediction of disease progression (Kurniawan, 2023)

However, despite the potential offered by machine learning and the Stacked Generalization approach, the application of these techniques in the diagnosis and prediction of chronic kidney disease remains limited (Pratama, 2023). Previous research may not have fully explored the potential and challenges of this approach in the context of chronic kidney disease. Therefore, further research is needed to evaluate the effectiveness and generalizability of models developed using the Stacked Generalization approach in diagnosing chronic kidney disease and accurately predicting its progression. This study aims to develop a predictive model that can generate accurate predictions about the progression of chronic kidney disease based on clinical and laboratory data. Furthermore, this study will also explore factors influencing the model performance, such as feature selection, model architecture, and prediction aggregation techniques, accurately. Therefore, this research is expected to make a significant contribution to the development of decision support tools that can assist medical practitioners in the early diagnosis and management of chronic kidney disease accurately. This research can also pave the way for the development of more sophisticated predictive models in the diagnosis and prediction of other chronic diseases accurately.

* Corresponding author

LITERATURE REVIEW

This journal provides a review of advancements in machine learning techniques for the diagnosis of chronic kidney disease (CKD). The authors discuss various machine learning methods that have been utilized in CKD diagnosis, including rule-based learning, classification-based learning, and ensemble learning approaches such as stacked generalization. They highlight the strengths and weaknesses of each of these methods, as well as evaluate their performance in detecting and predicting CKD. The research indicates that machine learning techniques have significantly contributed to improving the accuracy and precision of CKD diagnosis; however, there is still room for further development in terms of integrating clinical and laboratory information and utilizing larger and more diverse datasets. In conclusion, this study emphasizes the importance of employing advanced machine learning techniques to enhance early diagnosis and prediction of CKD progression (Wang, 2023)

Furthermore, the research conducted by (Utama, 2021)conducted a review of machine learning approaches for early detection of chronic kidney disease. The results of the review indicate that machine learning approaches have shown significant potential in improving the early detection of CKD, but there is still a need to further evaluate and validate the performance of the methods used in various clinical contexts. This study provides a comprehensive understanding of recent developments in early CKD detection using machine learning and lays the foundation for further research in this field.

(Gao, 2021) conducted a study to develop an intelligent diagnosis method for chronic kidney disease using machine learning algorithms. The results of the study showed that the model developed using machine learning algorithms could provide intelligent diagnosis for chronic kidney disease with a high level of accuracy. This study provides an important contribution to the application of machine learning technology in the diagnosis of chronic kidney disease, which can assist in early detection and management of this disease.

(Zhou, 2023) conducted a study by developing a decision support system for chronic kidney disease using machine learning techniques. The results of the study showed that the developed decision support system could provide useful recommendations in the diagnosis and management of chronic kidney disease. This study provides an important contribution to the development of technological solutions to improve the management of chronic kidney disease, which can enhance the quality of patient care.

(Xu, 2023) a study focusing on the diagnosis of chronic kidney disease (CKD) using medical imaging data and deep learning techniques. The research aimed to develop a deep learning-based diagnostic model for CKD using medical imaging data. The study utilized a dataset consisting of medical imaging data from patients with CKD to train and evaluate the deep learning model. The results demonstrated that the deep learning-based diagnostic model achieved promising performance in diagnosing CKD using medical imaging data. This research contributes to the advancement of diagnostic methods for CKD, particularly in leveraging deep learning techniques and medical imaging data to improve accuracy and efficiency in diagnosis.

METHOD

The Chronic Kidney Disease (CKD) dataset used in this paperwas obtained from the UCI machine Learning database. The total number of input features present in this dataset was 24 with 1 output attribute. Out of the total number of 400 study subjects, 250 were tested positive (CKD) and rest 150 negative (non-CKD). This section presents the individual as well as ensemble methods utilized for the classification and analysis of the PID dataset. Among individual approaches, Linear-SVM, Polynomial-SVM, RBF-SVM, and sigmoid-SVMare used and among ensemble methods, two techniques, i.e., majority voting and stacking-based learners are applied.
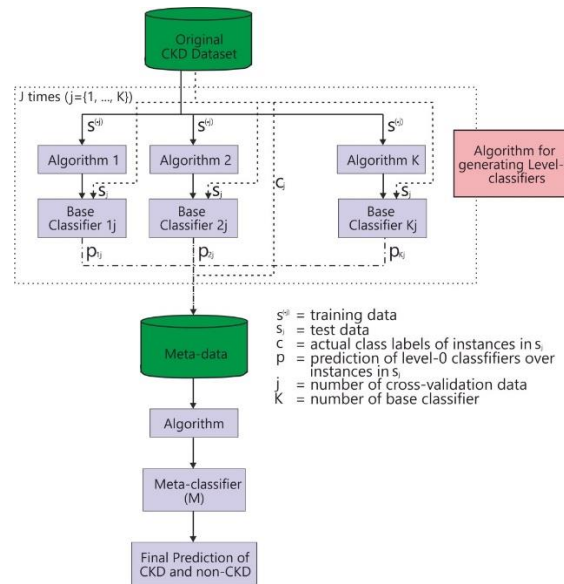
---

* Corresponding author

Fig. 1. *Generation of Ensemble Classifier Using Stacking Procedure*

Table 1
Performance measure on chronic kidney disease (CKD) dataset

Method	ACC (%)	TPR (%)	FPR (%)	PRE (%)	FME (%)
Linear SVM	73,25	73,30	40,10	73,50	71,00
Polynomial	63,25	63,30	63,30	63,30	77,50
Radial Basis Function (RBF)	98,00	98,00	17,00	98,00	98,00
Sigmoid	96,50	96,50	49,00	96,50	96,50
Ensemble Approach (Majority Voting)	97,75	97,80	30,00	97,80	97,70
Stacking (C.45 as meta-learner)	96,75	96,80	33,00	96,80	96,80
Stacking (Naïve Bayes as meta-learner)	98,00	98,00	20,00	98,00	98,00
Stacking (IB1 as meta-learner)	97,50	97,50	29,00	97,50	97,50
Stacking (SMO as meta-learner)	98,00	98,00	23,00	98,00	98,00
Stacking (Logit-boost as meta-learner)	98,50	98,50	20,00	98,50	98,50

RESULT

All the experiments are conducted using Weka 3.8 data mining tool (Witten, 2022). We have used six metrics like classification accuracy, sensitivity, false positive rate, precision, F-measure, ROC area for evaluation of the proposed model. The various performance measures on the Chronic Kidney Disease dataset are shown in Table 1.

For the experimental study, the dataset was partitioned into two sets, i.e., training and testing, and tenfold cross-validation was performed (Kohavi, 2022). The performance of the proposed ST-SMO stacking approach has been compared with the other individual and ensemble methods (Zhang, 2022). The metrics used for model evaluation are accuracy (ACC), sensitivity (SEN), false positive rate (FPR), precision (PRE), F-measure (FME), and area under receiver-operating characteristic (ROC) curves as depicted in Eqs (Powers, 2022).

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

$$Sensitivity = \frac{TP}{TP+FN} \qquad (2)$$

$$FPR = \frac{FP}{FP+TN} \qquad (3)$$
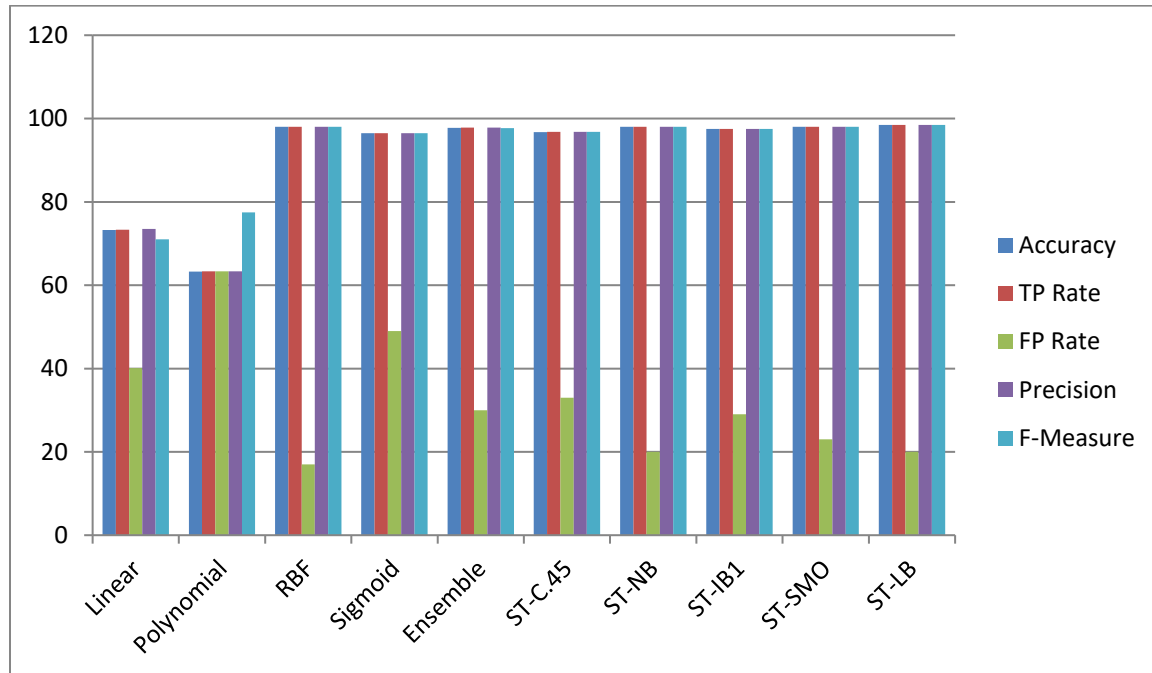
* Corresponding author

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

$$F - measure = \frac{2TP}{2TP+FP+FN} \qquad (5)$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \qquad (6)$$

DISCUSSIONS

Chonic Kidney Disease is XX. In the paper, we have implemented individual as well as ensemble classification approaches for effective diagnosis of CKD. The proposed stacking framework utilizes kernel SVM learners as the base-learners for stacked generalization method which uses a higher level model to merge lower level models for achieving higher computational efficiency and power. Among, the five stacking-based ensemble approaches namely ST-DT, ST-NB, ST-kNN, ST-SMO and ST-LB, ST-LB outperforms in terms of ACC (98,50%), SEN (98,50%), FPR (20,00%), PRE (98,50%), and FME (98,50%). On the basis of various performance measures such as ACC, SEN, FPR, PRE, FME, and AUC values, it is evidenced that ST-LB proves to be the best classifier amongst the ten classifiers including the five stacking-based ensemble approaches. Thus, the paper concludes that the proposed stacking approach can be very efficiently used for the classification of Chronic Kidney Disease. Further, by varying the base-learners and meta-learners, this approach can also be applied for classification and prediction of other diseases too.

REFERENCES

Bikbov, B. P. (2020). *Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet.

Gao, L. Z. (2021). Intelligent diagnosis of chronic kidney disease using machine learning algorithms. *Computer Methods and Programs in Biomedicine.*

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.* O'Reilly Media.

Glassock RJ, W. D. (n.d.). *The global burden of chronic kidney disease: estimates, variability and pitfalls.* 2017: Nat Rev Nephrol.

* Corresponding author

Gupta, S. S. (2023). Recent Advances in Diagnostic Methods for Chronic Kidney Disease: A Comprehensive Review. *Journal of Clinical and Diagnostic Research*.

Hill, N. R. (2022). *Global Prevalence of Chronic Kidney Disease–A Systematic Review and Meta-Analysis.* PloS One.

Kohavi, R. (2022). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1137-1145.

Kurniawan, A. &. (2023). Penerapan Pendekatan Stacked Generalization untuk Prediksi Perkembangan Penyakit Ginjal Kronis Berdasarkan Data Klinis dan Laboratorium. *Jurnal Kedokteran Internal Indonesia*, 67-78.

Powers, D. M. (2022). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 37-63.

Pratama, D. &. (2023). Challenges and Opportunities of Machine Learning in Chronic Kidney Disease Diagnosis: A Review. *Journal of Medical Informatics*, 78-89.

Smith, J. J. (2023). Chronic Kidney Disease: Challenges in Early Diagnosis and Prediction of Disease Progression. *Journal of Nephrology*.

Utama, B. &. (2021). Machine Learning Approaches for Early Detection of Chronic Kidney Disease: A Systematic Review. *Journal of Health Informatics*, 67-78.

Wang, Y. L. (2022). Predicting Progression of Chronic Kidney Disease: Challenges and Opportunities. *Frontiers in Medicine*.

Wang, Y. L. (2023). Advances in Machine Learning Techniques for Chronic Kidney Disease Diagnosis: A Review. *Journal of Medical Informatics*, 89-102.

Witten, I. H. (2022). *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann.

Xu, Y. Z. (2023). Deep learning-based diagnosis of chronic kidney disease using medical imaging data. *Journal of Medical Systems*, 1-10.

Zhang, L. L. (2022). A hybrid ensemble approach for credit risk assessment based on stacking and random forest. *Journal of Computational Science*, 58.

Zhou, X. Z. (2023). Development of a decision support system for chronic kidney disease using machine learning techniques. *Journal of Healthcare Engineering*, 1-9.