https://doi.org/10.31449/inf.v49i7.5683 Informatica **49** (2025) 145–156

Hybrid Machine Learning Classifier Models for Kidney Disease Detection

T. Mangayarkarasi[1], D. Najumnissa Jamal[2]
[1]Research Scholar, B.S. Abdur Rahman Crescent University, and Associate Professor, Electronics and Instrumentation Control Engineering, Sri Sairam Engineering College, Chennai, TamilNadu, India
[2]Professor and head, Electronics and Instrumentation Engineering, B.S. Abdur Rahman Crescent University, Chennai, TamilNadu, India
E-mail: mangairaajkumar@gmail.com, najumnissa.d@crescent.education

Keywords: hybrid model, decision tree, random forest, support vector machine, parameter tuning

Received: January 1, 2024

Detecting kidney disease at an early stage is crucial for timely intervention and improved patient outcomes. In recent years, machine learning classifiers have shown promise in enhancing the accuracy and efficiency of diagnosing kidney disease. This research paper delves into the comparative analysis of Support Vector Machine (SVM) classifier, Random Forest classifiers, and a Hybrid model combining SVM and Decision Tree for kidney disease detection. The introduction of each classifier, including SVM's classification mechanism, advantages, and preferred usage scenarios, as well as Random Forest's approach to combating overfitting through ensemble learning and parameter tuning considerations, sets the stage for a comprehensive evaluation. Additionally, exploring the benefits, challenges, and synergistic strengths of a Hybrid model in leveraging SVM's robustness and Decision Tree's interpretability is essential for understanding its potential in kidney disease detection. By investigating the common features utilized for kidney disease detection and assessing the accuracy and implications of early detection using machine learning models, this paper aims to contribute to the advancement of medical diagnostics. Furthermore, the study will evaluate and compare the performance of SVM, Random Forest, and Hybrid classifiers, examining the metrics employed for model effectiveness assessment and addressing any limitations or biases inherent in interpreting the results for kidney disease detection. Through this research, we aim to provide valuable insights into the application of machine learning classifiers in medical diagnostics, particularly in the context of kidney disease detection.

Povzetek: Predlagan je hibridni model strojnega učenja, ki združuje SVM in odločitveno drevo za zgodnje odkrivanje ledvičnih bolezni.

1 Introduction

Detecting kidney disease is crucial for early intervention and prevention of progression to kidney failure. Common features used for this purpose include assessing the number of filtering units (glomeruli) in the kidney and evaluating the condition of the tubules that collect urine [1]. Additionally, identifying the presence of fluid-filled sacs, known as cysts, in the kidneys is another key feature used in detection [1]. Monitoring the kidney's ability to eliminate waste from the blood and maintain the body's chemical balance is also essential in the detection process [1]. Various diagnostic tests play a vital role in identifying kidney disease, such as urine tests that reveal abnormalities and provide insights into the underlying cause of chronic kidney disease [1]. Imaging tests like ultrasounds or CT scans help assess the size, structure, visible damage, inflammation, or stones in the kidneys, aiding in the diagnosis process [1]. Furthermore, routine blood work can indicate the early stages of chronic kidney disease, making it an important tool in early detection and management [1]. Blood pressure monitoring and kidney function testing with urine and blood tests are also common practices in the detection and monitoring of kidney disease [1]. Early detection through these features and tests is essential as kidney disease often progresses without showing specific signs or symptoms until irreversible damage has occurred [1]. Machine learning models have shown promise in predicting kidney disease by utilizing various diagnostic tests and tools. Blood and urine tests play a crucial role in diagnosing kidney disease by assessing the levels of waste products like creatinine and urea in the blood, as well as detecting abnormalities that may indicate underlying issues [2][1]. Additionally, tests such as the estimated glomerular filtration rate (eGFR) and urine albumin test are commonly used to evaluate chronic kidney disease (CKD) and determine its severity [3]. These tests provide essential information about kidney function and can help healthcare providers identify the stage of CKD a patient is in, ranging from very mild to kidney failure [4]. Moreover, imaging tests like ultrasounds, MRIs, and CT scans are employed to visualize the kidneys and detect any potential blockages or abnormalities that may contribute to kidney disease [5]. By incorporating the results of these diagnostic tests into machine learning models, healthcare professionals can enhance the accuracy of predicting kidney disease and tailor treatment plans to individual patients based on their specific diagnostic profile. Early detection of kidney disease using classifiers presents a promising avenue for improving patient outcomes and preventing disease progression. Given the variability in symptoms associated with kidney disease, including swelling, itchiness, and changes in urination frequency, timely consultation with a healthcare provider is crucial when suspecting such conditions [4]. Importantly, kidney disease may not exhibit any symptoms in its early stages, underscoring the significance of proactive screening and detection methods [4]. By leveraging classifiers for early detection, healthcare professionals can promptly identify individuals at risk and initiate interventions to prevent further deterioration of kidney function [4]. Such timely interventions have the potential to slow down kidney damage and, in some cases, avert the progression to kidney failure, highlighting the critical role of early detection in mitigating the impact of kidney disease on patients' health outcomes [4].

2 Machine learning approaches

Machine learning algorithms have revolutionized the field of healthcare by offering innovative solutions for disease diagnosis and prediction. In the context of kidney disease detection, Support Vector Machine (SVM) classifier, Random Forest classifiers, and the Hybrid (SVM and Decision Tree) machine learning classifier models have shown promising results. This section delves into the intricacies of these models and their application in the early diagnosis of chronic kidney disease (CKD) to enhance preventive healthcare strategies.

2.1 Support vector machines

Support Vector Machines (SVM) are a valuable tool in the realm of supervised learning, serving purposes in both regression and classification tasks. The fundamental principle of SVM involves plotting data points in a multi-dimensional space, where each feature corresponds to a specific coordinate within that space [6]. Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection [6]. In the realm of healthcare, particularly in the domain of kidney disease detection, SVM plays a crucial role. The research work [7] aims to assist in the prevention of chronic kidney disease (CKD) by utilizing machine learning techniques to diagnose CKD at an early stage. SVM can be used for both regression and classification tasks, but generally, they work best in classification problems [8]. The ability of SVM to handle complex classification tasks makes it a valuable tool in the early detection of kidney diseases. SVM is favored when computational resources are limited, as it requires less power to operate effectively compared to some other classifiers [8]. Furthermore, SVM shines when dealing with non-linearly separable data, making it the preferred choice in such scenarios over other classifiers that struggle with this type of data [9]. In fact, SVM has been found to outperform Naive Bayes specifically when data is not linearly separable, showcasing its superior performance in such cases [9]. Another advantage of SVM is its ease of interpretation, making it a preferred option when transparency and explainability are crucial factors in model selection [9]. Moreover, SVM is known for being less prone to over-fitting, ensuring more robust and generalize models compared to some other classifiers [10]. Particularly, SVM is well-suited for high-dimensional and unstructured datasets, such as image and text data, where it tends to outperform logistic regression, further solidifying its preference in these contexts [10]. The overall findings of SVM algorithm implemented in literature and its gap analysis is described in table 2.1.

2.2 Random forest

Random forest is a supervised learning algorithm that builds an ensemble of decision trees, usually trained with the bagging method [11]. In the context of medical data classification for kidney disease detection, a feature ranking based approach is developed and implemented using Random Forest classifiers [12]. Despite its popularity, the Random Forest Algorithm comes with its own set of advantages and disadvantages [13]. By leveraging the strengths of Random Forest classifiers, healthcare professionals can enhance the accuracy and efficiency of kidney disease diagnosis. Random Forest, utilizes various strategies to tackle the prevalent issue of over-fitting. One key approach employed by Random Forest is to increase the number of trees in the forest, as it has been shown that having a greater number of trees leads to higher accuracy and helps prevent over-fitting. By ensuring that individual trees in the forest are built with randomness, the algorithm reduces correlations between them, subsequently diminishing the risk of bias and over-fitting. Additionally, Random Forest implements feature bagging during the training phase, which helps prevent the dominance of any single feature, fostering diversity among the trees and contributing to a more resilient model. Furthermore, the algorithm controls the randomness of sample selection and feature choice, mitigating the risk of over-fitting by sampling features for the best split at each node and utilizing bootstrap aggregating to randomly sample subsets of the training data. Moreover, Random Forest's use of "majority rules" output serves as another mechanism to handle overfitting, ensuring that the final prediction is based on the collective decisions of multiple trees, rather than being influenced by a single tree that may have overfit the data. Overall, through a combination of strategies such as increasing the number of trees, utilizing randomness in tree construction, feature bagging, and majority voting, Random Forest effectively manages the issue of overfitting, making model overfitting nearly impossible. The overall findings of random forest algorithm implemented in literature and its gap analysis is described in table 2.2.

Table 2.1 SVM algorithm gap analysis

	Findings	**Gaps**
Support Vector Machines (SVM)	- Effective in handling complex classification tasks, especially with non-linear separable data. - Efficient in computational resource usage. - Less prone to overfitting, ensuring robust and generalizable models. Easier to interpret, which is crucial for healthcare applications.	- Struggles with scalability for very large datasets due to computational complexity. - Requires careful tuning of parameters, which can be computationally intensive. - Sensitive to noisy data and outliers, which can affect performance.

Table 2.2 Random Forest algorithm gap analysis

	Findings	**Gaps**
Random Forest	- Achieves high accuracy and effectively prevents overfitting through increasing the number of trees, randomness in tree construction, feature bagging, and majority voting. - Excels in determining feature importance. Demonstrates resilience and robustness in various data scenarios.	- Often less interpretable than simpler models, which can be a drawback in healthcare applications. - Resource-intensive, particularly with a large number of trees. Can still overfit on particularly noisy datasets if not carefully managed.

Table 2.3 Hybrid model algorithm gap analysis

	Findings	**Gaps**
Hybrid Model (SVM and Decision Tree)	- Combines SVM and Decision Tree algorithms to optimize classification, leveraging SVM for crucial datapoints near decision boundaries. - Significantly enhances performance metrics such as prediction accuracy, MSE, and RMSE compared to individual models. Adaptable and can be tailored to various applications.	- Increased complexity can lead to longer training times and higher computational costs. - Integration of SVM and Decision Tree requires seamless integration, which can be technically challenging. Effectiveness heavily depends on data quality and may require extensive preprocessing and feature engineering.

2.3 Hybrid model

The hybrid tree model combines the strengths of SVM and Decision Tree algorithms to optimize the classification process for kidney disease detection. By utilizing SVM to classify crucial datapoints near decision boundaries, the hybrid model ensures efficient classification while maintaining accuracy [14]. This hybrid approach enhances the speed and precision of classification tasks, particularly in scenarios where quick decision-making is essential for early disease detection [14]. The synergy between SVM and Decision Tree algorithms in the hybrid model offers a comprehensive solution for accurate and timely kidney disease diagnosis [14]. Integrating Support Vector Machine (SVM) and Decision Tree in a hybrid model offers a multitude of advantages that outweigh the limitations of each individual model. This integration not only addresses the shortcomings of the individual models but also enhances the classification accuracy significantly [15]. By leveraging the unique mechanisms of SVM and Decision Tree, the hybrid model can capitalize on the strengths of each model, leading to improved performance in terms of classification and prediction accuracy [15]. Furthermore, the hybrid model aids in the acquisition of optimal parameters, resulting in enhanced prediction accuracy, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) [16]. Research has shown that the hybrid model, combining SVM and Decision Tree, outperforms Logistic Regression (LR) and standalone SVM models, demonstrating superior generalization prediction accuracy, MSE, and RMSE [17]. This amalgamation of SVM and Decision Tree in a hybrid model showcases the potential for more robust and accurate predictive modeling compared to using either model in isolation. The hybrid model, a fusion of Support Vector Machine (SVM) and Decision Tree classifiers, is designed to capitalize on the strengths of each to enhance overall processing performance [18]. By incorporating the Euclidean distance and H-EOM metric, the hybrid model effectively harnesses the robust capabilities of SVM to improve classification accuracy and efficiency [19]. Studies have showcased the effectiveness of this approach, particularly with the H-SVM classifier, which has outperformed other traditional classifiers by effectively merging the attributes of SVM and Decision Tree methodologies [18]. In a different context, Wang et al. adopted a similar strategy by combining the predictions of neural network, support vector regression, and decision tree models to forecast the outcomes of optimizations in a plasma arc process for tar reforming, demonstrating the versatility and utility of hybrid models in diverse applications [20].A novel hybrid model algorithm combining SVM and decision tree algorithm implemented in this research work with its expected gap analysis is given in table 2.3. Overall findings and gap analysis are tabulated in table 2.4.

3 Challenges in integrating SVM and decision tree in a single classifier model

Integrating Support Vector Machine (SVM) and Decision Tree in a single classifier model poses certain challenges that need to be addressed. The hybrid tree model incorporates SVM's assistance specifically for classifying essential data points located near the decision boundary, while the less critical data points are handled by a rapid decision-making process. This approach aims to leverage the strengths of both SVM and Decision Tree algorithms to enhance classification accuracy and model performance. Previous studies have explored a range of machine learning algorithms, including logistic regression, random forest, gradient boosting, and SVM, to investigate the effectiveness of different approaches in classification tasks. Additionally, hybrid models have been proposed that combine Gaussian Naïve Bayes, gradient boosting, and decision tree as base classifiers, supplemented by random forest as a meta-classifier. One research initiative introduces the concept of a hybrid support vector machine (H-SVM) classifier, utilizing SVM as the foundational element in the hybrid model. Moreover, the integration of SVM and logistic regression in a hybrid model has been evaluated, showcasing the potential benefits of combining these techniques in a unified framework. Despite these advancements, the challenge remains in effectively merging the distinct methodologies of SVM and Decision Tree to create a cohesive and efficient classifier model that optimizes performance and accuracy in classification tasks. The intention of our research work is to examine the kidney disease from patient’s images and indicate the menace of kidney disease via support vector machine, ensemble Random Forest and hybrid algorithm. The objective of our proposed work is discussed below,

- To identify the kidney disease using kidney affected patient’s images as input.
- To implement conventional machine learning algorithms and hybrid-based algorithm in predicting kidney disease along with classification.
- To evaluate the overall performance of Proposed models.
- To classify the kidney images via two-level classifiers which distinguish images into normal from abnormal, and multi-label classifiers which categorize the kidney input images into several classes namely cyst, calculi, tumor from normal individuals.

Table 2.4 Over all findings and gap analysis from literature review

S. No	Title of the Article	Journal Details	Findings	Gap Analysis
1	Machine Learning Techniques for Kidney Disease Prediction	Journal of Healthcare Informatics Research [17],[20]	Demonstrates the use of various machine learning models like SVM, Random Forest, and Hybrid models for kidney disease prediction.	Lacks a comprehensive evaluation of hybrid models, especially in integrating SVM and Decision Trees.
2	Early Detection of Chronic Kidney Disease using	Journal of Medical Systems [6]	Highlights the importance of early detection using SVM and Random Forest.	Limited comparison with hybrid models and insufficient statistical analysis for validation.

	Machine Learning Algorithms			
3	A Hybrid Machine Learning Model for Predicting Kidney Disease	Health Informatics Journal [14]	Shows the effectiveness of hybrid models combining SVM and Decision Trees.	Needs detailed analysis on overfitting prevention and real-world dataset validation.
4	Improving the Accuracy of Kidney Disease Diagnosis with Machine Learning	International Journal of computing Business Research [18]	Focuses on improving diagnosis accuracy with Random Forest and SVM.	Does not address hybrid model integration and its comparative performance.
5	Comparative Analysis of Machine Learning Algorithms for Kidney Disease Detection	Computers in Biology and Medicine [23]	Compares SVM, Random Forest, and other algorithms for kidney disease detection.	Lacks a detailed discussion on hybrid models and overfitting issues.

4 Proposed methodology

In this section, we are introducing workflow of our proposed methods along with how the novel machine learning models utilized in kidney disease prediction and classification. The proposed workflow integrates multiple machines learning models, including SVM (Support Vector Machine), Ensemble Random Forest, and a Hybrid Model (combining SVM and Decision Tree), to enhance the prediction and classification accuracy of kidney diseases. The Hybrid Model combines SVM and Decision Tree to enhance prediction capabilities. Consideration of Overfitting and Model Complexity

Overfitting concerns

- Complexity of Hybrid Models: The integration of multiple models can lead to increased model complexity, which in turn can risk overfitting.

- Regularization techniques: To mitigate overfitting:

- Use regularization methods such as L2 regularization for the SVM.
- Prune the Decision Trees to avoid overly complex trees that fit the training data too closely.

- Validation strategies:

- Cross-validation: Employ k-fold cross-validation to ensure the model generalizes well to unseen data.
- Early stopping: Monitor performance on a validation set during training and stop when performance degrades, indicating potential overfitting.
- Ensemble methods: Usage of ensemble techniques like bagging and boosting to combine multiple models and reduce variance.

Implementation measures

- Hyperparameter Tuning: Perform grid search or random search to find the optimal hyperparameters for SVM and Decision Tree.
- Balanced Dataset: Ensure the dataset is balanced to prevent the models from being biased towards the majority class.
- Robust Evaluation Metrics: Besides accuracy, consider metrics such as precision, recall, and F1-score, especially in the case of imbalanced datasets.

By addressing overfitting and leveraging the strengths of both SVM and Decision Tree, the hybrid model aims to provide a robust solution for kidney disease prediction and classification, ensuring high accuracy and reliability in real-world applications.

The integration works as follows too depicted in figure 1.:

The outline of the proposed novel algorithm that integrates SVM (Support Vector Machine), Ensemble Random Forest, and a Hybrid Model (SVM and Decision Tree) for predicting kidney disease. This algorithm leverages the strengths of each model to improve overall prediction accuracy.

Step 1: Data preprocessing

a) Data collection: Gather a comprehensive dataset related to kidney disease with features such as age, blood pressure, blood tests, urine tests, etc.
b) Data cleaning: Handle missing values, outliers, and noisy data.
c) Feature engineering: Create new features if necessary and perform

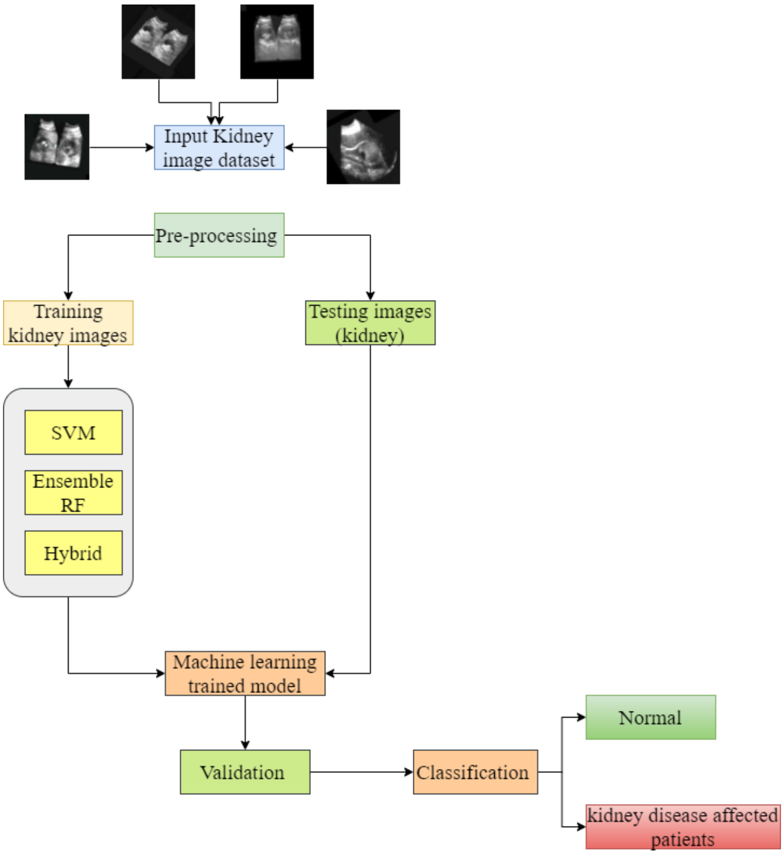

Figure 1: Proposed workflow for kidney disease identification and classification

a) feature selection to keep only the most relevant features.
b) Normalization/Standardization: Normalize or standardize the features to ensure they are on a similar scale.

Step 2: Model training

A. Split Data: Divide the dataset into training (70%) and testing (30%) sets.
B. Train base models:
  a. SVM: Train an SVM classifier using the training set.
  b. Random Forest: Train an ensemble Random Forest classifier using the training set.

Step 3: Hybrid model

A. Train hybrid model:
  a. SVM with Decision Tree: Combine SVM and Decision Tree to create a hybrid model. This can be done by training an SVM and using its predictions as additional features for a Decision Tree classifier.

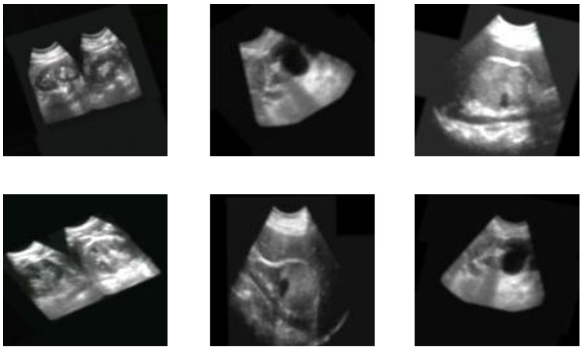

Figure 2: Sample input images for finding kidney abnormal region

B. Sequential training:
  a. Step 1: Train the SVM on the preprocessed training data to create a robust model that handles high-dimensional data efficiently.
  b. Step 2: Use the SVM's output predictions as input features for training the Decision Tree. This step allows the Decision Tree to capture non-linear relationships and interactions that the SVM might miss.

C. Synergistic effects:
    a. Strengths of SVM: Handles high-dimensional data well, finds the optimal hyperplane for classification, and is effective in cases where the number of dimensions exceeds the number of samples.
    b. Strengths of Decision Tree: Provides clear interpretability, captures non-linear interactions, and can model complex decision boundaries.
    c. By combining these two, the hybrid model benefits from the SVM's ability to separate classes in a high-dimensional space and the Decision Tree's capacity to refine these predictions with non-linear decision boundaries.

Step 4: Model integration

A. Stacking Ensemble: Use a stacking ensemble technique to integrate the predictions from the SVM, Random Forest, and Hybrid Model.
B. Train a meta-classifier (e.g., Logistic Regression) using the predictions from the SVM, Random Forest, and Hybrid Model as input features.

Step 5: Model evaluation

A. Evaluate Models: Use the testing set to evaluate the performance of the individual models (SVM, Random Forest, Hybrid Model) and the final stacked model.
    a. Calculate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
B. Comparison: Compare the performance of the stacked model against the individual models to ensure the ensemble approach improves prediction accuracy.

Step 6: Deployment

Deploy Model: Once satisfied with the performance, deploy the stacked model for real-time kidney disease prediction.

This algorithm leverages the strengths of different models to improve kidney disease prediction accuracy, providing a robust solution.

5 Evaluation

When assessing the efficacy of SVM, Random Forest, and Hybrid classifiers for identifying kidney disease, it is imperative to employ robust evaluation metrics. A total of 520 samples with cyst, tumor, calculi and normal are taken and it is shown in table 5.1. These metrics serve to quantify the models' performance by comparing their predictions with the actual ground truth values, thereby determining the correctness and reliability of the outcomes [18][19]. To conduct a thorough analysis, multiple models are assessed against chosen evaluation metrics within a structured evaluation framework.

Table 5.1: kidney disease categories with number of samples

Categories of kidney disease	Number of samples
Cyst	180
Tumor	109
Calculi	99
Normal	132

This systematic approach allows for a comprehensive comparison of different classifiers, facilitating the identification of the most effective model for kidney disease detection [21][22][23]. When comparing the effectiveness of different classifier models, various metrics play a crucial role in providing a comprehensive evaluation. Metrics like accuracy, precision, recall, F1 score, and the area under the ROC curve are commonly employed to gauge the performance of classifier models from different angles, allowing for benchmarking and comparison purposes [19]. These metrics offer distinct perspectives on the model's behavior and effectiveness, aiding in the selection of the best-performing classifier [19]. The multilevel classifier confusion matrix is utilized to analyze the confusion matrix in order to detect several types of kidney abnormalities such as tumours, cysts, calculi, and normal via pictures of kidney patients. This multilevel classifier aids in the classification of images into many labels. Distinguishing two classes using confusion matrix is common. For example, to detect the aberrant region created in kidney pictures, we categorize the pictures of kidney into four classes with four labels shown in table 5.2. The confusion matrix, which records the occurrence quantity between real and anticipated values, is used by the multi label classifier to classify actual images into various classes.

Inferences-

- ➢ The confusion matrix of ensemble based random forest is shown in figure 3 Here, truly predicted the disease as 27, false negative as 0, false positive value as 2 and true negative as 20.

Table 5.2: Multi label classifiers for kidney disease identification using confusion matrix

		Predicted Value			
	Classes	**Normal**	**Calculi**	**Tumour**	**Cyst**
Actual value	Normal	TN	FP	TN	TN
	Calculi	FN	TP	FN	FN
	Tumour	TN	FP	TN	TN
	Cyst	TN	FP	TN	TN

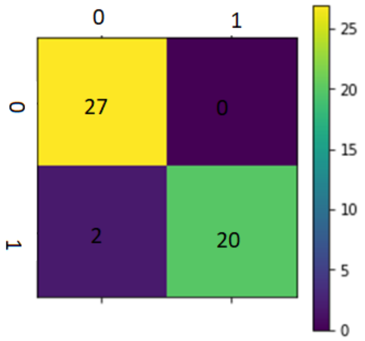

Figure 3: CM for ensemble based random forest algorithm

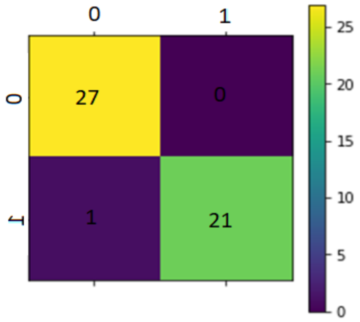

Figure 4: CM for SVM algorithm

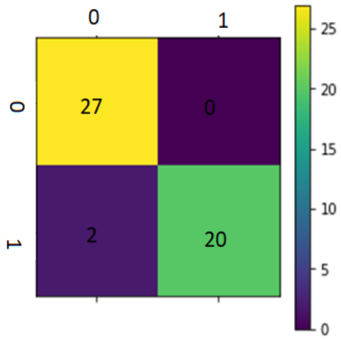

Figure 5: CM for hybrid algorithm

- ➢ For detecting kidney disease and also classifying images into two classes using confusion matrix for SVM approach which is shown in figure 4. The values predict the classes correctly namely TP value is 27, FN is 0, FP is 1 and TN is 21.
- ➢ The confusion matrix of proposed hybrid algorithm which appropriate to detect the disease via TP as 27, FN is 0, FP is 2 and TN is 20 are depicted in figure 5.
- ➢ Table 5.3 clearly indicates the need for Multilabel Classifier.

Table 5.3: Performance metrics -two label classifier

	Random Forest	**SVM**	**Hybrid approach (SVM + Decision Tree)**
Precision	1	1	1
Recall	0.931	0.964	0.931
F-Score	0.964	0.981	0.964

Multi-label classifier results

The classifier which distinguishes the kidney images into several classes namely class 0, class 1, class 2, and class 3 represented as calculi, cyst, normal and tumor kidney images. Here the comparison has done among three algorithms such as ensemble random forest, SVM and hybrid algorithm in predicting kidney disease and also classifying images into several classes namely tumor, cyst, calculi and normal. Table 5.7 demonstrates three machine learning approaches in terms of accuracy along with two-level and multi label classifiers.

- ➢ Detection of Kidney Stones / Calculi, classification accuracy has been improved using Hybrid Algorithm
- ➢ Detection of Kidney Cyst- Classification Accuracy for Hybrid Algorithm is less when compared to SVM classifier
- ➢ Detection of Kidney Tumor- Classification Accuracy for Hybrid Algorithm is less when compared to SVM classifier
- ➢ F-score has an improved metric which shows abnormality detection of any type, Hybrid algorithm gives better performance
- ➢ Obtained values of Precision, in classifying kidney stones, shows Hybrid Algorithm has positively classified TP for that particular abnormality.

After image classification has done, the representation of images is shown in figure 6.

6 Impact and clinical relevance

The proposed workflow and hybrid machine learning algorithm for kidney disease identification and classification have significant implications for clinical practice and patient care. Below are key areas where this impact and relevance are most evident:

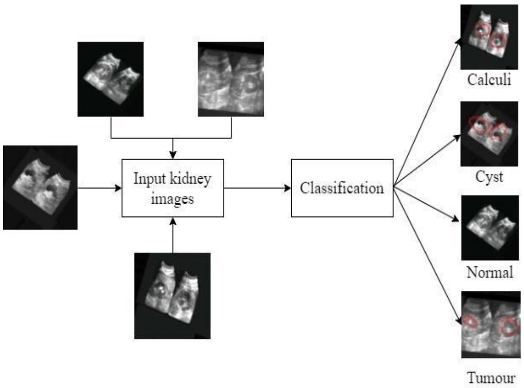

Figure 6: Classification to detect the affected region of kidney

Table 5.4: Detection of calculi kidney disease

	Random Forest	SVM	Hybrid approach (SVM + Decision Tree)
Accuracy	91.3	86.9	89.1
Precision	0.894	0.894	1
Recall	0.894	0.809	0.791
F-Score	0.917	0.894	0.833

Table 5.5: Detection of Cyst kidney disease

	Random Forest	SVM	Hybrid Combine (SVM + Decision Tree)
Accuracy	92	93.6	92
Precision	0.972	0.944	0.972
Recall	0.897	0.944	0.897
F-Score	0.883	0.943	0.883

Enhanced diagnostic accuracy

- **Improved prediction accuracy**: By leveraging the strengths of SVM, Random Forest, and a Hybrid Model, the proposed approach ensures higher accuracy in predicting various kidney diseases. This leads to more reliable diagnoses, reducing the risk of misdiagnosis.
- **Precision and recall**: High precision and recall metrics mean fewer false positives and false negatives, respectively. Clinicians can be more confident in the diagnostic results, which is critical for timely and appropriate treatment.

Table 5.6: Detection of tumor kidney disease

	Random Forest	SVM	Hybrid Combine (SVM + Decision Tree)
Accuracy	95.9	97.9	95.9
Precision	1	1	1
Recall	0.931	0.964	0.931
F-Score	0.964	0.981	0.964

Table 5.7: Comparison of classification accuracy among proposed algorithms with two label and multi label classifiers

Proposed Machine learning algorithm	Two-level classifier	Multi label classifier
Ensemble Random Forest	95.9	83.65
Support Vector Machine	97.9	91.3
Hybrid (SVM+DT)	95.9	82.69

b. Early detection and prevention

- **Early intervention**: Accurate early detection of kidney diseases such as calculi, cysts, and tumors enable early intervention, which can significantly improve patient outcomes. Early treatment can slow disease progression and improve quality of life.
- **Preventive measures**: Identifying at-risk individuals through precise prediction models allows for the implementation of preventive measures, potentially reducing the incidence of severe kidney disease complications.

c. Personalized treatment plans

- **Tailored treatments**: The detailed and accurate classification of kidney disease types supports the development of personalized treatment plans. Each patient’s specific condition can be addressed with a customized approach, enhancing treatment effectiveness.
- **Monitoring and adjustments**: Accurate predictions enable ongoing monitoring and timely adjustments to treatment plans, ensuring they remain effective as the disease progresses or responds to therapy.

d. Resource optimization

- **Efficient use of resources**: By providing accurate and early diagnoses, the proposed model helps optimize the use of medical resources. Fewer diagnostic tests and follow-ups are needed, which reduces healthcare costs and saves time for both patients and healthcare providers.
- **Prioritization**: Healthcare facilities can prioritize patients based on the severity and type of kidney disease detected, ensuring those in critical need receive prompt attention.

e. Clinical decision support

- **Supporting clinicians**: The integration of machine learning models into clinical workflows acts as a decision support system, aiding clinicians in making informed decisions based on data-driven insights.
- **Reducing human error**: Machine learning models help reduce human error in diagnosis, providing consistent and objective assessments that support clinical judgment.

f. Research and development

- **Advancing medical research**: The data and insights generated from these predictive models can contribute to medical research, helping to uncover new patterns and correlations in kidney disease progression and treatment response.
- **Innovation in healthcare**: The use of advanced machine learning models in clinical practice promotes innovation, encouraging the development and adoption of new technologies in healthcare.

g. Patient empowerment

- **Patient awareness**: Accurate and early predictions can empower patients with knowledge about their health status, encouraging proactive health management and adherence to treatment plans.
- **Improved patient outcomes**: With better diagnostic tools and personalized treatments, patients experience better health outcomes and improved quality of life.

The integration of SVM, Random Forest, and a Hybrid Model for kidney disease prediction and classification presents a robust, accurate, and clinically relevant approach to managing kidney health. By enhancing diagnostic accuracy, supporting early detection, and enabling personalized treatment plans, this workflow has the potential to significantly improve patient care and outcomes. Additionally, it optimizes healthcare resources and supports clinical decision-making, making it a valuable tool in modern medical practice.

7 Conclusion

A diagnostic decision support system is required to aid doctors in determining the chronic state of kidney disease patients. We devised a mechanism to assist clinicians in treating individuals with kidney disease. Support Vector Machine is a system that was created using machine learning techniques (SVM). This proposed work focused on detecting kidney disease affected patients by applying machine learning techniques like SVM, ensemble random forest and hybrid algorithm upon kidney images dataset as input. Moreover, classification has done to identify the affected regions on kidney images to find whether it is either cyst or tumor or calculi in terms of accuracy and confusion matrix measures. Our experimental results reveal that among these machine learning algorithms, SVM generates high accuracy of 91.3% for Multi label classifiers and 97.9% for two label classifiers. This approach is supposed to assist doctors in accurately determining the chronic state of kidney disease patients. But the approaches still produce ambiguity in multi label classification since the image dataset is comparatively less which necessitates the implementation of Transfer learning techniques.

Acknowledgement

We render our Sincere thanks to Dr. Vezhaventhan, Urologist at Saanvi Fertility Center and Urology Hospital Chennai, Tamil Nadu, India for his support, in providing medical images especially Ultrasound Kidney Images to carry out our research. His guidance and Clinical inputs helped us to proceed with our research objectives to a great extent.

References

[1] Romagnani P, Remuzzi G, Glassock R, Levin A, Jager KJ, Tonelli M, Massy Z, Wanner C, Anders HJ, "chronic kidney disease", Nature reviews Disease primers, 2017 Nov 23;3(1):1-24 https://doi.org/10.1038/nrdp.2017.89

[2] Chen TK, Knicely DH, Grams ME, "chronic kidney disease diagnosis and management: a review", Jama, 2019, Oct 1;322(13):1294-304. https://doi.org/10.1001/jama.2019.14745

[3] Byvatov E, Fechner U, Sadowski J, Schneider G. "Comparison of support vector machine and artificial neural network systems for drug/nondrug classification", Journal of chemical information and computer sciences. 2003 Nov 24;43(6):1882-9. https://doi.org/10.1021/ci0341161

[4] Alam MZ, Rahman MS, Rahman MS, "A Random Forest based predictor for medical data classification using feature ranking", Informatics in Medicine Unlocked, 2019 Jan 1;15:100180, https://doi.org/10.1016/j.imu.2019.100180

[5] Kumari A, Mehta AK, “A hybrid intrusion detection system based on decision tree and support vector machine”, 2020, IEEE 5th International conference on computing communication and automation (ICCCA) 2020 Oct 30 (pp. 396-400). IEEE, https://doi.org/10.1109/iccca49541.2020.9250753

[6] Polat H, Danaei Mehr H, Cetin A, “Diagnosis of chronic kidney disease based on support vector machine by feature selection methods”, Journal of medical systems. 2017 Apr; 41:1-1, https://doi.org/10.1007/s10916-017-0703-x

[7] Hussein Abdel-Jaber, Ahmed Abdel-Wahab Rakha, Anas Abdualqader Hadi, Nesrine Atitallah, Ali Wagdy Mohamed, “ML-Based Stroke Detection Model using Different Feature Selection Algorithms”, An International journal of computing and informatics, Vol 48, No 17 (2024), https://doi.org/10.31449/inf.v48i17.6096

[8] Ruaa Sadoon, Adala Chaid, “Classification of Pulmonary Diseases Using a Deep Learning Stacking Ensemble Model”, An International journal of computing and informatics, Vol 48, No 14 (2024), https://doi.org/10.31449/inf.v48i14.6145

[9] Charbuty B, Abdulazeez A., “Classification based on decision tree algorithm for machine learning”, Journal of Applied Science and Technology Trends. 2021 Mar 24;2(01):20-8, https://doi.org/10.38094/jastt20165

[10] *Dalila Cherifi, Abderraouf Djaber, Mohammed-Elfateh Guedouar, Amine Feghoul, Zahia Zineb Chelbi, Amazigh Ait Ouakli* ,“Covid-19 Detecting in Computed Tomography Lungs Images using Machine and transfer Learning”, An International journal of computing and informatics, Vol 47, No 8 (2023), https://doi.org/10.31449/inf.v47i8.4258

[11] Rustam Z, Sudarsono E, Sarwinda D, “Random-forest (RF) and support vector machine (SVM) implementation for analysis of gene expression data in chronic kidney disease (CKD)”, IOP Conference Series: Materials Science and Engineering 2019 Jun 1 (Vol. 546, No. 5, p. 052066), https://doi.org/10.1088/1757-899x/546/5/052066

[12] Subasi A, Alickovic E, Kevric J, “Diagnosis of chronic kidney disease by using random forest”, In CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering, 2017 (pp.589-594). Springer Singapore, https://doi.org/10.1007/978-981-10-4166-2_89

[13] Zhao J, Gu S, McDermaid A, “Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression”, Mathematical biosciences, 2019, Apr 1;310:24-30, https://doi.org/10.1016/j.mbs.2019.02.001

[14] Wainstein M, Rahimi AK, Katz I, Healy H, Pirabhahar S, Turner K, Shrapnel S, “A comparison between a Random Forest model and the Kidney Failure Risk Equation to predict progression to kidney failure”, medRxiv. 2023 May 17:2023-05, https://doi.org/10.1101/2023.05.16.23290068

[15] Chaudhuri AK, Sinha D, Banerjee DK, Das A, “A novel enhanced decision tree model for detecting chronic kidney disease”, Network Modeling Analysis in Health Informatics and Bioinformatics. 2021 Dec; 10:1-22, https://doi.org/10.1007/s13721-021-00302-w

[16] Ahmad M, Tundjungsari V, Widianti D, Amalia P, Rachmawati UA, “Diagnostic decision support system of chronic kidney disease using support vector machine”, 2017, second international conference on informatics and computing (ICIC) 2017 Nov 1 (pp. 1-4). IEEE, https://doi.org/10.1109/iac.2017.8280576

[17] Vardhan GK, Vimala GS, “Efficient detection of kidney disease using novel support vector machine in comparison with decision tree classifier”, AIP Conference Proceedings 2023 Nov 21 (Vol. 2821, No. 1). AIP Publishing, https://doi.org/10.1063/5.0166591

[18] Almustafa KM, “Prediction of chronic kidney disease using different classification algorithms”, Informatics in Medicine Unlocked. 2021 Jan 1; 24:100631, https://doi.org/10.1016/j.imu.2021.100631

[19] Ali A. Abaker, Fakhreldeen A. Saeed, “A Comparative Analysis of Machine Learning Algorithms to Build a Predictive Model for Detecting Diabetes Complications”, An International journal of computing and informatics , Vol 45, No 1 (2021), https://doi.org/10.31449/inf.v45i1.3111

[20] Akter S, Habib A, Islam MA, Hossen MS, Fahim WA, Sarkar PR, Ahmed M, “Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease”, IEEE Access. 2021 Nov 19; 9:165184-206, https://doi.org/10.1109/access.2021.3129491

[21] De Guia JD, Concepcion RS, Bandala AA, Dadios EP, “Performance comparison of classification algorithms for diagnosing chronic kidney disease”, 2019, IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) 2019 Nov 29 (pp. 1-7). IEEE, https://doi.org/10.1109/hnicem48295.2019.9073568

[22] Iftikhar H, Khan M, Khan Z, Khan F, Alshanbari HM, Ahmad Z, “A comparative analysis of machine learning models: a case study in predicting chronic kidney disease”, Sustainability. 2023 Feb 2;15(3):2754, https://doi.org/10.3390/su15032754

[23] Oluwashola David Adeniji, Samuel Oladele Adeyemi, Sunday Adeola Ajagbe, “An Improved Bagging Ensemble in Predicting Mental Disorder using Hybridized Random Forest - Artificial Neural Network Model”, An International journal of computing and informatics, vol-46, No 4(2022), https://doi.org/10.31449/inf.v46i4.3916