Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Comparing Machine Learning Techniques for Detecting Chronic Kidney Disease in Early Stage

Md Abdur Rakib Rahat[1], MD Tanvir Islam[2], Duc M Cao[3], Maliha Tayaba[4], Bishnu Padh Ghosh[5], Eftekhar Hossain Ayon[6] ✉ Nur Nobe[7], Taslima Akter[8], Mamunur Rahman[9] and Mohammad Shafiquzzaman Bhuiyan[10]

[1]*Department of Electrical and Computer Engineering (ECE), North South University, Dhaka, Bangladesh*
[2]*Department of Computer Science, Monroe College, New Rochelle, New York, USA*
[3]*Department of Economics, University of Tennessee, Knoxville, TN, USA*
[4]*Department of Computer Science, University of South Dakota, Vermillion, South Dakota, USA*
[5]*School of Business, International American University, Angeles, California, USA*
[6]*Department of Computer & Info Science, Gannon University, Erie, Pennsylvania, USA*
[7]*Department of Healthcare Management, Saint Francis College, Brooklyn, New York, USA.*
[89]*Department of Science in Information Technology, Washington University of Science & Technology, USA*
[10]*Department of Business Administration, Westcliff University, Irvine, California.*

Corresponding Author: Eftekhar Hossain Ayon, **E-mail**: ayon001@gannon.edu

| ABSTRACT

In medical care, side effect trial and error processes are utilized for the discovery of hidden reasons for ailments and the determination of conditions. In our exploration, we used a crossbreed strategy to refine our optimal model, improving the Pearson relationship for highlight choice purposes. The underlying stage included the choice of ideal models through a careful survey of the current writing. Hence, our proposed half-and-half model incorporated a blend of these models. The base classifiers utilized included XGBoost, Arbitrary Woods, Strategic Relapse, AdaBoost, and the Crossover model classifiers, while the Meta classifier was the Irregular Timberland classifier. The essential target of this examination was to evaluate the best AI grouping techniques and decide the best classifier concerning accuracy. This approach resolved the issue of overfitting and accomplished the most elevated level of exactness. The essential focal point of the assessment was precision, and we introduced a far-reaching examination of the significant writing in even configuration. To carry out our methodology, we used four top-performing AI models and fostered another model named "half and half," utilizing the UCI Persistent Kidney Disappointment dataset for prescient purposes. In our experiment, we found out that the AI model XGBoost classifier gains almost 94% accuracy, a random forest gains 93% accuracy, Logistic Regression about 90% accuracy, AdaBoost gains 91% accuracy, and our proposed new model named hybrid gains the highest 95% accuracy, and performance of Hybrid model is best on this equivalent dataset. Various noticeable AI models have been utilized to foresee the event of persistent kidney disappointment (CKF). These models incorporate Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine, K-nearest neighbor, LDA (Linear Discriminant Analysis), GB (Gradient Boosting), and neural networks. In our examination, we explicitly used XGBoost, AdaBoost, Logistic Regression, Random Forest, and Hybrid models with the equivalent dataset of highlights to analyze their accuracy scores.

| KEYWORDS

Glomerular Filtration Rate, liver cancer, chronic kidney failure

| ARTICLE INFORMATION

ACCEPTED: 15 December 2023 **PUBLISHED:** 01 January 2024 **DOI:** 10.32996/jcsts.2024.6.1.3

1. Introduction

Persistent kidney sickness (CKD) positions as the sixteenth most noteworthy reason for long stretches of life lost on a worldwide scale. Successful screening, determination, and the executives completed by essential consideration experts are fundamental in turning away negative results related to CKD, for example, cardiovascular sickness, end-stage kidney illness, and mortality. Persistent kidney sickness (CKD) is a typical condition, with an expected predominance of 13% among grown-ups, in light of the

Wellbeing Study for Britain in 2009. The gamble of CKD ascends with age, and it frequently co-happens with hypertension, diabetes, and cardiovascular illness (CVD). CKD is much of the time ignored as it regularly needs unambiguous side effects. A critical minority of CKD patients will advance to end-stage kidney sickness, and postponed commencement of renal substitution treatment in this gathering can prompt expanded grimness and mortality. Notwithstanding, the main part of CKD is its job as a free, strong, and modifiable gamble factor for CVD. CKD is additionally emphatically connected to other huge antagonistic results, including intense kidney injury, feebleness, and expanded mortality. Constant kidney sickness (CKD) is a huge general wellbeing worry, with expanding frequency and predominance. It influences roughly 10-13% of the populace and is a main source of significant mortality and bleakness around the world. Deplorably, both public and clinical familiarity with CKD stays deficient. Persistent kidney illness (CKD) is a condition portrayed by kidney brokenness or primary irregularities. It is regularly characterized as having a glomerular filtration rate (GFR) of under 60 mL/min/1.73 m^2 or the presence of kidney marker harm. This harm might incorporate signs like albuminuria, unusual pee silt, electrolyte uneven characters, and different issues connected with rounded messes. Primary irregularities can likewise be distinguished through histology or imaging, and a background marked by kidney transplantation enduring no less than 90 days. These rules are as per the KDIGO rules. Significantly, CKD can be analyzed without essentially recognizing its particular reason.

The gamble of creating ongoing kidney illness (CKD) ascends with age, and it frequently exists together with conditions like hypertension, diabetes mellitus, cardiovascular infections (CVD), obstructive rest apnea, or episodes of intense kidney injury. Besides, metabolic elements, particularly insulin obstruction, dyslipidemia, and hyperuricemia, have been connected to the beginning and movement of CKD. A few examinations recommend that CKD is more common in men. Quite often, African Americans are at a higher gamble of kidney harm compared with Caucasians. Strangely, there's proof to propose that low birth weight could make people more vulnerable to kidney infection. Also, smoking, over-the-top liquor utilization, and sporting medication use have been related to the movement of CKD.

Nonetheless, a portion of these gamble variables can be changed, possibly postponing or, in any event, forestalling the headway of kidney disappointment. It is critical to think about fitting therapy for ongoing kidney illness (CKD) in all patients, no matter what their CKD stage. Numerous scientists advocate for the early assessment and treatment of CKD patients to dial back their movement. This approach isn't just financially savvy but additionally brings about decreased patient grimness and further developed results, including easing back the progression of kidney disappointment and diminishing the gamble of cardiovascular illness (CVD). Also, it's fundamental to address comorbid conditions. The essential results of CKD incorporate advancing to kidney disappointment, confusion emerging from declining kidney capability, and the improvement of CVD. Expanding proof proposes that early recognition and treatment can forestall or defer a portion of these unfriendly results. Screening is suggested for high-risk populations, which include people with hypertension, diabetes mellitus, and those aged north of 65.

This screening includes pulse observation, urinalysis, and the estimation of serum creatinine to appraise GFR. We want to give an outline of ongoing headways in how we might interpret the sub-atomic systems hidden in constant kidney sickness. We have focused on subjects, for example, oxidative pressure, the contribution of fiery cells, neutrophil gelatinase-related lipocalin, lattice metalloproteinase, the stomach kidney hub, and arising treatment focuses that have become exposed as our insight into these instruments has extended. Persistent kidney sickness (CKD) presents a growing medical services challenge. In 2009-2010, CKD consumed 1.3% of the all-out NHS financial plan, with over portion of this consumption coordinated toward the 2% of people who advance to kidney disappointment. The assessed cost, credited to the abundance of instances of strokes and respiratory failures in that year contrasted with an age and orientation-matched populace without CKD, went from £174 to £178 million.

Persistent kidney illness (CKD) is characterized as the presence of kidney harm or an expected glomerular filtration rate (eGFR) under 60 ml/min/1.73 m^2, continuing for something like 3 months, no matter what the hidden reason. It addresses a dynamic decrease in kidney capability, at last prompting the requirement for renal substitution treatment, which can be as dialysis or transplantation. Ongoing kidney disappointment is a pervasive kind of kidney problem that happens when both kidneys endure harm, prompting determined well-being challenges for those distressed. Kidney harm envelops any condition that impedes the organ's legitimate working, whether brought about by a basic sickness or a lack of basic variables like the Glomerular Filtration Rate (GFR). Our created expectation model uses clinical side effects as information and utilizes a stacking classifier in AI, with the Irregular Backwoods ML model as the basic classifier, to produce forecasts. The use of AI in medical services determination is progressively imperative because of its ability to direct high-level examinations, diminish the potential for human mistakes, and upgrade expectation precision. AI models and classifiers are, as of now generally recognized as exceptionally functional instruments for the ID of different ailments, including cardiovascular breakdown, diabetes, growths, and liver disease, among others. Inside the clinical field, an assortment of AI calculations are utilized for both grouping and relapse undertakings. These calculations envelop

Guileless Bayes, SVM, XGBoost, AdaBoost, [Khan et al. 2023, Miah et al. 2023, Kayyum et al. 2020, Islam et al. 2020, Miah et al. 2022] and choice trees for arrangement, while strategic relapse, irregular woodland, and direct relapse are utilized for relapse purposes. The effective usage of these calculations works with the beginning phase identification of sicknesses, consequently lessening death rates and empowering brief therapy for patients. Besides, people with ongoing kidney disappointment are encouraged to consolidate proactive tasks like activity, keep up with satisfactory water admission, and stay away from the utilization of undesirable food varieties in their day-to-day schedules, as well as observe their clinical side effects.

2. Related Work

Rahman et al. [2019] research scientists have acquainted different methodologies to guarantee that patients have the chance to get the necessary clinical treatment, fully intent on lessening the movement of the illness or possibly accomplishing a fix. Recognizing persistent kidney sickness (CKD) at the beginning phase is of the most extreme significance. To accomplish this, a model was created by using digitized ECG information obtained from straightforwardly accessible data sets, incorporating PTB for people with renal circumstances and Capriccio from the Physio Net Data set for those without such circumstances. The model's legitimacy was, in this way, affirmed by utilizing extra information from a similar data set, effectively sorting people as either sound or distressed with CKD. It was seen that while consolidating the qualities of the QT and RR spans, the model displayed a precision level of 97.6%, outperforming the exactness accomplished while using either highlight in detachment.

Chittora et al. [2021] research outcomes show that while utilizing the complete Engineered Minority Over-Examining Procedure (Destroyed), the Straight Help Vector Machine (LSVM) with an L2 punishment accomplishes the most noteworthy exactness level at 98.86%. The going chart represents the assessment and examination of different measurements, for example, review, region under the bend, F-measure, accuracy, and the GINI coefficient, for various strategies. After carrying out the completely unlocked Destroyed strategy, the techniques, including the Determination of Administrator Relapse and Least Outright Shrinkage, reliably yielded the best results. Also, inside the Engineered Minority Over-testing Method structure, which utilized the most un-outright Shrinkage and chosen highlights by the determination administrator, the Direct Help Vector Machine exhibited the most elevated precision at 98.46%. Moreover, notwithstanding the AI models, the equivalent dataset went through examination by a profound brain organization.

Nikhila et al.'s [2021] study report, the creators have assessed the adequacy of four group calculations in distinguishing proof of early ongoing kidney sickness in patients. They evaluated the exhibition of these AI calculations utilizing seven key execution measures, including Exactness, Particularity, Awareness, F1-Score, and the Matthew Connection Coefficient. The examination uncovered that AdaBoost and Arbitrary Woodland outflanked Angle Supporting and Stowing about exactness, accuracy, and awareness. Besides, AdaBoost and Irregular Backwoods accomplished an ideal 100 percent exactness for both the Matthew Relationship Coefficient and the Region under the Curve. These proposed AI calculations can enable clinical experts in the early location of constant kidney sickness, offering a trustworthy strategy for forestalling the beginning of the disease.

In the Ekanayake et al. [2020] paper, the creators carried out a work process that pointed toward using clinical information to foresee the seriousness of persistent kidney sickness (CKD). This interaction includes quality determination, cooperative separating for overseeing missing qualities, and information pre-handling. Among the 11 AI methods explored, the gradual tree classifier and irregular woodland classifier showed the most elevated exactness and the least characteristic inclination. This study features the significance of consolidating space ability while utilizing AI calculations for CKD expectation, as well as tending to useful contemplations in information assortment.

Nandhini et al. [2021] studied the examination, which included anticipating results utilizing four unique grouping calculations: Arbitrary Woods Classifier, Strategic Relapse, K-Closest Neighbor (KNN), and Backing Vector Machine (SVM). These calculations were applied to 400 datasets from the UCI store, each containing 25 credits. The examination discoveries show that KNN, LR, and SVM accomplished exactness levels of 94%, 98%, and 93.75%, separately. Interestingly, the Irregular Woods (RF) classifier accomplished the most noteworthy exactness level, arriving at an ideal 100 percent.

The study by Jhumka et al. [2022] aims to anticipate the visualization of constant kidney sickness (CKD) by taking into account different variables. The review utilized a freely accessible dataset containing data gathered in India. To set up the dataset for examination, a few procedures were utilized to address missing qualities and exceptions. Accordingly, the dataset was separated into CKD and non-CKD cases utilizing both Irregular Woods and Profound Brain Organizations (DNN). While looking at the consequences of these two techniques, the DNN model accomplished a higher double grouping exactness of 98.8%.

Desai et al. [2019] accentuate the significance of early location and anticipation of Constant Kidney Sickness (CKD) utilizing Information Mining methods. By dissecting patient records, the Boruta investigation Information Mining calculation can foresee the probability of creating CKD by thinking about different elements, including factual, verifiable, and clinical data. The dataset for this study was acquired from UCI and comprises 400 information tests from the southern piece of India, covering people aged 2-90 years. The calculation's assessment of the meaning of these elements can give a gauge of CKD events, which could demonstrate the importance of fostering the sickness to people in danger. Besides, the utilization of the Boruta Investigation strategy makes the analysis more productive and open for patients as it is uninhibitedly accessible. Given the extended expense of $12 billion for treating current and new instances of kidney disappointment in Australia by 2020, the expected advantages of this calculation become much more critical, highlighting the significant significance of early location and avoidance of CKD.

Ghosh et al. [2020] research specialists utilized four particular AI strategies, specifically AdaBoost, Backing Vector Machine with Edge Amplification (SMM), Straight Discriminant Investigation (LDA), and Angle Helping, to anticipate results precisely. They tried these calculations utilizing a dataset obtained from the UCI AI Vault. The outcomes showed that Slope Helping accomplished the most elevated precision, with an amazing score of 99.80%. Moreover, the review included a scope of execution assessment measures to survey the viability of these calculations, giving a premise for choosing the most reasonable calculation for the given errand.

All the work is done for specific datasets, but no one gives a sustainable model that can explore the model's performance on a more diverse set of datasets. However, none of the work conducts comparative studies with other contemporary models to assess the model's competitiveness. So, in our work, we focused on these two issues to get attention for this health issue.

3. Methodology

Constant kidney sickness is a moderate, hopeless condition related to huge disease and mortality, common among grown-ups, especially those with diabetes and hypertension. Protecting kidney capability is significant and can be accomplished through non-pharmacological methodologies, like dietary and way-of-life adjustments, as well as designated pharmacological therapies for constant kidney illness. A plant-based, low-protein, low-salt eating routine might assist with relieving glomerular hyperfiltration and support kidney capability for longer, possibly prompting ideal changes in corrosive base equilibrium and the stomach microbiome. Drugs that influence intracranial hemodynamics, for example, modulators of the renin-angiotensin-aldosterone pathway and SGLT2 (SLC5A2) [Islam et al. 2020, Aljaaf et al. 2018, Salekin et al. 2016] inhibitors, can safeguard kidney capability by bringing down intraglomerular pressure autonomously of pulse and glucose control. Novel specialists, such as non-steroidal mineralocorticoid receptor bad guys, could shield the kidney through calming or ant-fibrotic instruments. Certain kidney infections, both glomerular and cystic, may profit from sickness-explicit treatments. Dealing with the cardiovascular gamble related to persistent kidney illness, limiting disease risk, and forestalling intense kidney injury are essential measures for these patients, given the significant entanglements and the job of unusual gamble factors in this condition. At the point when renal substitution treatment becomes vital, a progressive change to dialysis can be thought of, possibly protecting the remaining kidney capability longer. There are similitudes and contrasts between kidney-safeguarding care and strong consideration. Further investigation into dietary and drug intercessions, as well as inventive techniques, is fundamental to guarantee ideal kidney-saving consideration and to upgrade the life span and personal satisfaction for impacted people.

We planned a half-and-half model utilizing different Python libraries, including Pandas, Scikit-Learn, Matplotlib, and Plotly. This model was tried utilizing the Persistent Kidney Disappointment (CKF) dataset [Khan et al. 2022, Yildirim 2017] acquired from the UCI storehouse, which contains data connected with kidney disappointment classified into two gatherings: CKF (addressed as 1) and non-CKF (addressed as 0) [Emon et al. 2021, Khan et al. 2023]. To guarantee the consistency and reproducibility of our outcomes, we chose the AI calculation with the most noteworthy precision for our examination and execution. Likewise, our crossover model integrates numerous calculations, including Gaussian Credulous Bayes (GNB), Gullible Bayes (NB), and Choice Tree (DT) [Haque et al. 2023, Jonayet et al. 2023, Shukla et al. 2023] as base classifiers, alongside Arbitrary Woods as the Meta classifier. Our model was created in view of the experiences acquired during assessment and execution. We embraced a tree-based AI calculation to accomplish high exactness and address overfitting concerns. To distinguish exceptions in our dataset, we used a violin plot, as shown in Figures 1 and 2. To alleviate the gamble of overfitting, we executed the k-overlap cross-approval calculation and customized our model appropriately. In the resulting segments, we offer a thorough conversation of every one of these classifiers.

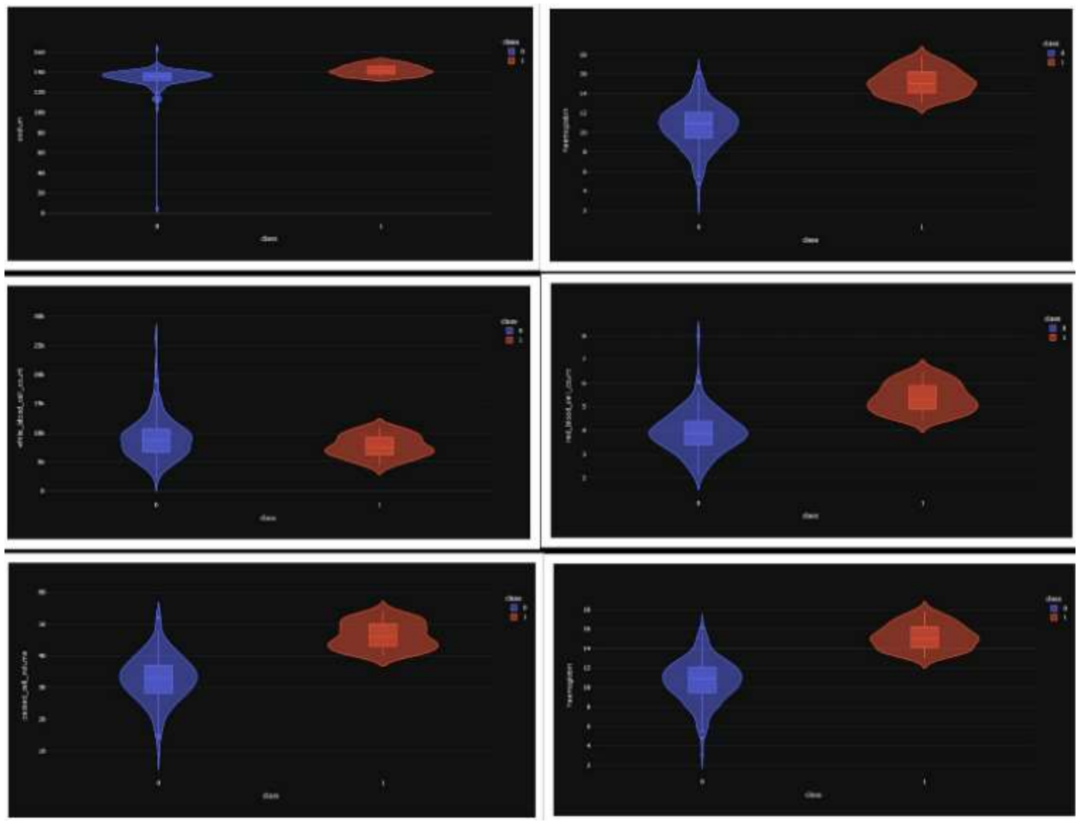

Figure 1: Violin plot of attributes

3.1 XGBoost

XGBoost, short for Outrageous Slope Supporting, is a notable AI calculation eminent for its surprising prescient capacities. This calculation works by making a group of choice trees and using inclination supporting, a strategy that iteratively improves model precision by revising blunders. XGBoost is a flexible instrument equipped for dealing with different errands, including order, relapse, and positioning. It offers the adaptability to redo misfortune capabilities to meet explicit prerequisites. XGBoost additionally integrates regularization methods to forestall overfitting and upholds custom misfortune capabilities for different undertakings, alongside giving highlights to dissecting the significance of individual elements. One of the vital qualities of XGBoost is its high productivity. It upholds equal handling, making it computationally productive, and is accessible in different programming dialects. Because of its flexibility and execution, XGBoost is generally utilized in information science rivalries and genuine applications, making it a favored decision for information researchers and simulated intelligence experts.

3.2 Random Forest (RF)

The administered learning strategy incorporates the notable Arbitrary Timberland AI calculation, which can be applied to both order and relapse issues in AI. It depends on the idea of gathering realization, where different classifiers are joined to resolve complex issues and improve the model's presentation. In an irregular woodland, each tree is developed utilizing an alternate subset of preparing information and an irregular subset of info highlights. This approach further develops execution, lessens overfitting, and upgrades the model's capacity to sum up.

The irregular timberland calculation likewise utilizes a method called packing, which includes haphazardly inspecting the preparation information with substitution to improve the model's precision. Utilizing the chosen subset of preparing information and highlights, the calculation assembles different choice trees, each prepared to anticipate the objective variable [Alshakrani et al. 2021, Dissanayake et al. 2023]. During expectation, the irregular timberland calculation makes a last forecast by consolidating the results of the multitude of individual trees. For order purposes, it returns the class with the largest number of votes from each tree, while for relapse, it gives the normal of each tree's expectations.

3.3 Logistic Regression

Strategic relapse is a managed learning calculation essentially utilized for paired grouping undertakings, where the objective is to anticipate the likelihood of a case having a place with one of two classes, commonly alluded to as "achievement" and "not achievement." This likelihood is addressed by Condition 1. Calculated relapse accomplishes this by applying a straight model to the info highlights and going the result through a sigmoid capability, which scales the result values somewhere in the range of 0 and 1. During the preparation cycle, the calculation changes the model's loads and inclination terms to limit a particular misfortune capability, frequently the strategic misfortune. One of the remarkable benefits of calculated relapse is its interpretability; you can undoubtedly grasp what each component means for the order result by inspecting the coefficients [19,20].

Strategic relapse tracks down applications in different fields, including medication, money, and normal language handling, where twofold or multi-class arrangement undertakings are common.

Condition 1:
P = 1/(1 + e^(- (b0 + b1x + b2x^2)))

In Condition 1, "P" addresses the likelihood of a case having a place with the "achievement" class, while "x" means the info highlights, and "b0," "b1," and "b2" are the model boundaries to be gotten the hang of during the preparation cycle. The sigmoid capability guarantees that the anticipated likelihood falls inside the scope of 0 to 1 [Aruna et al. 2023, Kashyap et al. 2022].

3.4 AdaBoost

AdaBoost [Rajeshwari et al. 2022], short for Versatile Helping, is a troupe AI calculation essentially used for arrangement undertakings. It works by making areas of strength for a model through the mix of numerous powerless students, frequently straightforward choice trees or "stumps." AdaBoost allocates higher loads to the misclassified information in every emphasis, permitting resulting powerless students to focus on correcting these blunders. The last characterization choice is reached by amassing the weighted forecasts of the multitude of powerless students. This iterative interaction goes on until a predefined number of powerless students are prepared or an ideal degree of exactness is accomplished.

AdaBoost is famous for its vigor, its capacity to deal with loud information, and its ability to make precise classifiers. These characteristics have made it a well-known decision in different applications, including PC vision and discourse acknowledgment, where characterization undertakings are normal.

3.5 Hybrid Model

A half-breed model calculation is a strategy that consolidates at least two unique models to resolve complex issues or make precise forecasts. The crucial thought behind a half-and-half model is to use the qualities of different models while moderating their singular shortcomings, bringing about better execution and more exact results. Crossover models find normal applications in fields like AI, information mining, and computerized reasoning.

One illustration of a crossbreed model includes joining a brain network with a choice tree. In this arrangement, the brain network is answerable for distinguishing examples and connections inside the information, while the choice tree is utilized to decipher these examples and make expectations in light of the learned standards. This blend improves the model's accuracy and gives a complete comprehension of how expectations are produced. The execution of a half-and-half model normally includes the accompanying advances:

1. Preprocessing: The information is preprocessed to clean and change it into a configuration reasonable for different procedures.
2. Feature Selection: Pertinent elements are looked over the information to decrease dimensionality and work on the model's effectiveness.
3. Ensemble Construction: Different AI strategies are coordinated to make the cross breed model. This combination can happen through strategies like mixing, stacking, or flowing.
4. Training: The cross breed model is prepared utilizing the preprocessed information and chosen highlights.
5. Prediction: The half breed model is utilized to make expectations in light of new information by consolidating the results of the various strategies.

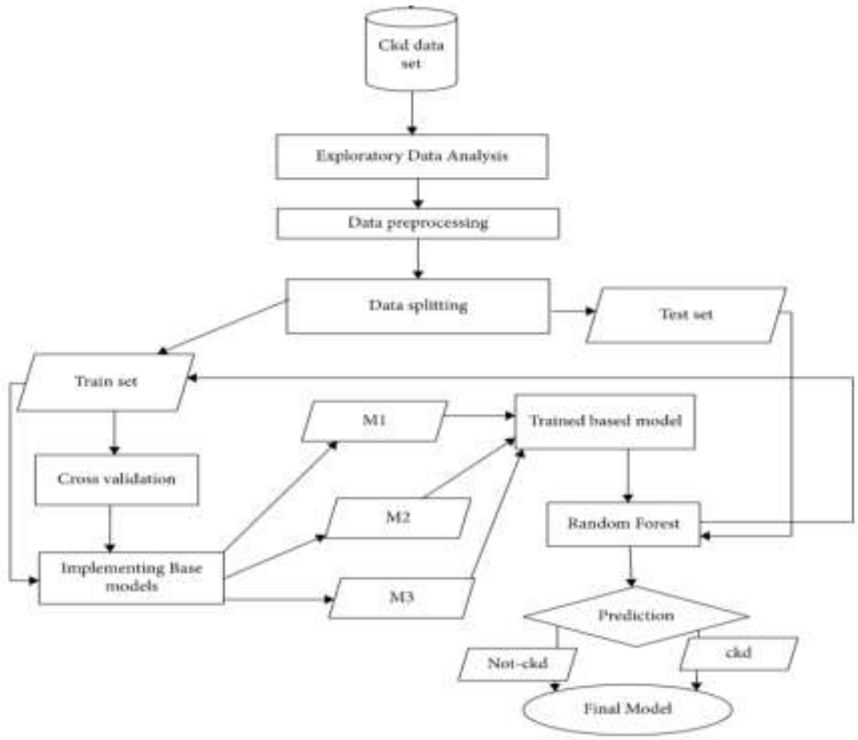

Figure 2: The overview of the study

By consolidating the qualities of different models and following an organized cycle, crossover models can offer improved execution and bits of knowledge, making them significant devices in different spaces. Without a doubt, the essential goal of a half-and-half model calculation is to make another model that outperforms the singular methods utilized in seclusion concerning exactness, heartiness, and generalizability. By combining different strategies, half-and-half models can catch a more extensive range of examples inside the information, relieve the gamble of overfitting, and eventually upgrade the general presentation of the model. This approach uses the correlative qualities of various techniques, bringing about additional exact and flexible answers for complex issues.

In our examination, we utilized the UCI Constant Kidney Disappointment (CKF) dataset, which we partitioned into two sections. To prepare the AI calculation, we used 80% of the information. We utilized the Irregular Woodland calculation as well as a mix of four different calculations: Mixture Model, Gaussian Guileless Bayes, Gullible Bayes, and Choice Tree [Snegha et al. 2020, Aljaaf et al. 2018]. We held the leftover 20% of the information for testing, during which we applied these calculations to anticipate constant kidney illness. We then, at that point, imagined and contrasted the anticipated qualities and the genuine qualities. Our methodology offers a few key benefits:

(I) We saddled the force of four unmistakable AI calculations [Emon et al. 2021], including Choice Tree, Credulous Bayes, Gaussian Guileless Bayes, and Arbitrary Woods. Besides, we joined these models utilizing stacking classifiers to make a half-and-half model.

(ii) We assessed the exhibition and exactness of these calculations on the equivalent dataset and contrasted their scores with recognizing the best-performing model.

(iii) By applying the stacking classifier strategy, we had the option to make another model with further developed exactness, improving the nature of our expectations.

This procedure exhibits a deliberate way of dealing with the model turn of events and testing, planning to accomplish the most elevated prescient precision for ongoing kidney sickness recognition.

4. Details of the Dataset

In our concentration on constant kidney disappointment, we chose a sum of fourteen (14) credits from the UCI store dataset to act as info highlights. These traits incorporate patient age, pulse (BP), pee explicit gravity, pee aluminum level, irregular blood glucose, glucose level, blood urea level, blood sodium level, serum creatinine level, blood potassium level, hemoglobin level, pressed cell volume, white platelet (WBC) count, and red platelet (RBC) count. We utilized information representation procedures to comprehend the organization of the CKF dataset, including the dispersion of patients with ongoing kidney disappointment and those without. A histogram plot, as portrayed in Figure 3, represented this. Sound cases were addressed by 0.0, while patients with ongoing kidney disappointment were addressed by 1.0. As per the dataset, there were 150 solid people and 250 people with constant kidney infections.

Figure 3: Histogram plot

To choose the most important highlights for distinguishing persistent kidney disappointment, you utilized the Pearson connection property determination strategy. Figure 5 shows the connections between the result name and the 14 information credits.

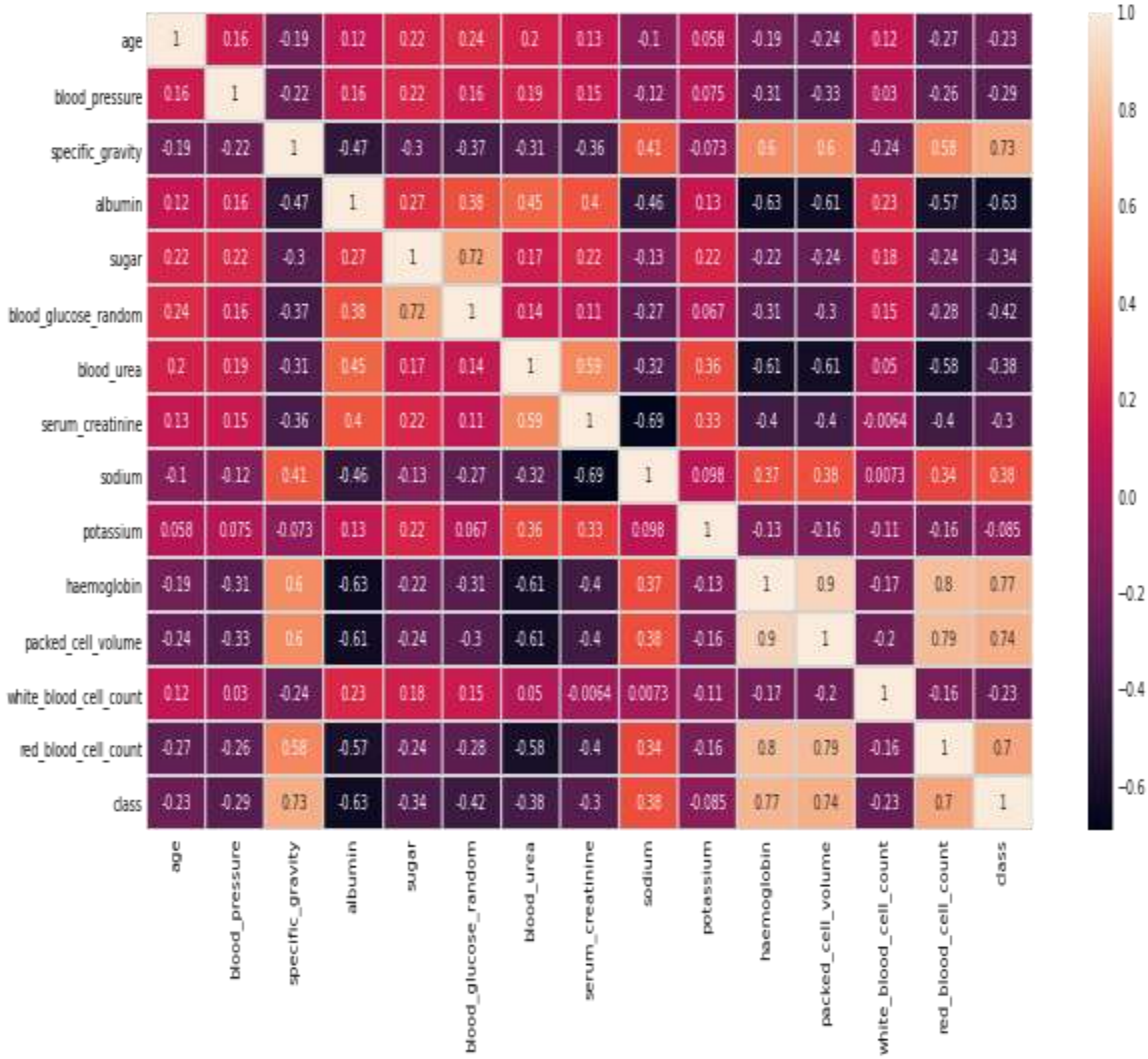

Figure 4: Heat map of chosen attributes

We likewise progressed from exploratory information examination to showing pair plots, which can uncover relationships between highlights, both consistent and straight out. The Sea born library was utilized to create these plots, giving an exhaustive and outwardly engaging portrayal of information. In your exploratory information examination, you utilized violin plots to picture the dispersion and thickness of the properties utilized in your half-breed model. Violin plots give a definite perspective on the dataset, showing the total information dispersion for every one of the 14 credits, as represented in Figure 2. This approach can be especially helpful for acquiring experiences into the information's qualities and dispersion.

5. Result and Discussion

To conjecture the event of constant kidney failure, we used a scope of AI calculations, including the Hybrid Model, XGBoost, AdaBoost, Logistic Regression, and a Random Forest classifier. The proposed approach includes utilizing a hybrid model that amalgamates both techniques to resolve the issue of overfitting and improve boost accuracy. In the production of the suggested hybrid model, specialists joined the boost accuracy of these calculations with the overfitting issue. The outcomes exhibited that the suggested model, outperforming the singular accuracy of every algorithm, accomplished a surprisingly close to 95% accuracy rate. Besides, we noticed a decrease in the gamble of overfitting with the usage of the hybrid model.

A provincial-level review was directed to survey the upgrades made in prediction models for ongoing renal failure. Most of the examination in this field has been led in less evolved nations because of their higher powerlessness to persistent renal failure, as made sense of in the presentation. Table 1 offers an exhaustive rundown of the investigation of different qualities utilizing various algorithms.

Table I: Analysis of Different Machine Learning Models

Models	Accuracy (%)	Precision (%)	Recall (%)	F-1 Score (%)	AUROC (%)
XGBoost	94	92.56	93.42	95.2	94.1
Random Forest	93	90.4	93	91.3	90.7
Logistic Regression	90	92	92.7	92	91.32
AdaBoost	91	91	90.5	94.5	90.79
Hybrid Model	95	96.2	95.1	96	94.56

The table represents the performance metrics of a few AI models in a particular task, including Precision, Accuracy, Review, F-1 Score, and AUROC, all communicated in rate values. These measurements give important experiences into the models' capabilities:

Among the AI models analyzed the hybrid Model stands apart with remarkable execution across all assessment standards. It achieves an impeccable accuracy of close to 95%, guaranteeing exact expectations for all cases. It accomplishes a high accuracy of 96.2%, demonstrating a significant extent of accurately anticipated positive occasions. Besides, the Hybrid Model succeeds in review, precisely recognizing 95.1% of genuine positive examples. These strong accuracy and review values result in a great F1-score of 96%, a well-balanced model. Moreover, the Hybrid Model achieves an AUROC score of 94.56%, highlighting its capacity to separate between positive and negative examples. While XGBoost comes close with a 94% accuracy, review and F1-score are marginally lower contrasted with our Mixture Model. Logistic Regression and AdaBoost exhibit decent execution; however linger behind in unambiguous regions, for example, accuracy and F1-score. Random Forest (RF) performs adequately, by and large; however, its measurements miss the mark regarding those of the half and half Model and XGBoost. In synopsis, the Mixture Model displays prevalent execution across all perspectives, settling on it as a convincing decision for precise and trustworthy expectations. Graph 1 represents the viability of different AI calculations.

At last, the decision of the most appropriate model ought to be dependent upon the particular requests and compromises of the undertaking. The Hybrid Model succeeds in general execution; however, the setting of the issue and the overall significance of precision and recall ought to direct the model choice cycle.

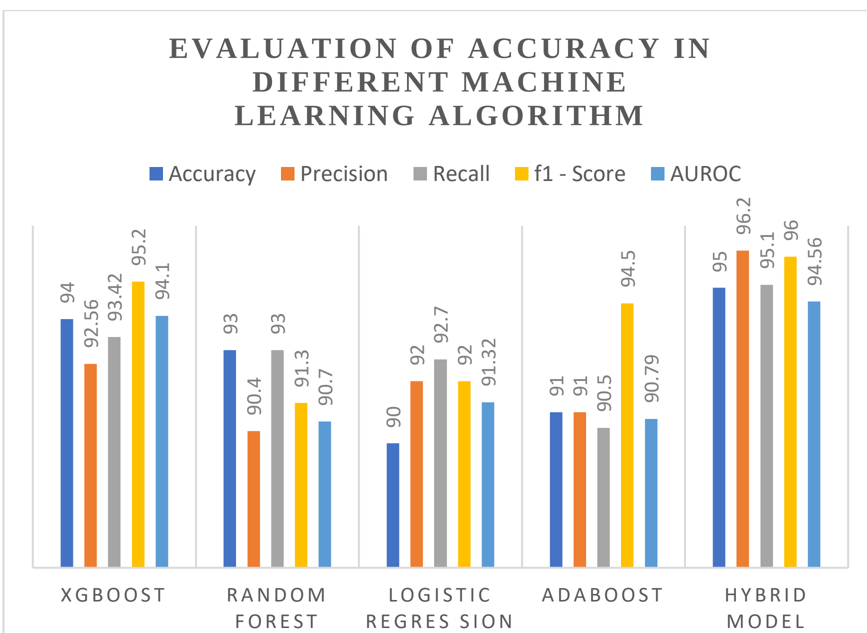

Chart 1: Accuracy level of different model

6. Conclusion and Future Work

The hybrid model, which combines the strengths of different algorithms, emerged as the top performer with exceptional accuracy, precision, recall, F1-score, and AUROC values. It achieved nearly 95% accuracy, demonstrating its ability to make highly accurate predictions. Moreover, it displayed a remarkable precision of 96.2%, indicating the model's proficiency in correctly classifying positive cases. Its recall of 95.1% showed that it effectively identified true positive instances. The F1 score of 96% signifies a well-balanced model that excels in both precision and recall. The high AUROC score of 94.56% underscores its capacity to distinguish between positive and negative cases. While the hybrid model stood out, other machine learning algorithms, such as XGBoost, Random Forest, Logistic Regression, and AdaBoost, also displayed respectable performances. However, the choice of the most suitable model should depend on the specific requirements and trade-offs of the task. Overall, the study provides a comprehensive evaluation of machine learning models for CKD prediction, highlighting the potential of hybrid models to enhance accuracy and the importance of selecting the right model for the specific problem. This research has the potential to contribute to the early detection and management of CKD, ultimately improving the quality of healthcare for affected individuals.

In less evolved nations, serious kidney disappointment is a predominant and extreme medical problem frequently connected to an absence of active work. Albeit different indicative strategies have been utilized by specialists, AI has arisen as a later viable methodology, flaunting better exactness than other existing techniques. This review puts specific accentuation on AI procedures and uses the Pearson relationship-based highlight choice technique to improve the accuracy of the AI classifier. The review directed cross-approval to assess the precision of the hidden classifiers, including XGBoost, AdaBoost, Strategic Relapse, and Arbitrary Backwoods [Arif-Ul-Islam et al. 2019, Arafat et al. 2021, Abu et al. 2023], all utilizing the equivalent dataset. Moreover, a dataset comprising 400 instances of persistent kidney disappointment (CKF), involving 14 unmistakable highlights obtained from the UCI information base, was used.

By applying the Pearson Connection Coefficient strategy for highlight choice and utilizing a stacking calculation with the best artificial intelligence classifiers for preparation, our proposed model accomplished close to 100% exactness in distinguishing the presence of CKF in people. To additional upgrade the presentation of the stacking model, a cross-approval technique was utilized. This stacking technique showed ideal execution while managing parallel CKF information. We accept that this stacking approach can be applied in foreseeing different illnesses and accomplishing considerably more significant levels of exactness. Overall, the study provides a comprehensive evaluation of machine learning models for CKD prediction, highlighting the potential of hybrid models to enhance accuracy and the importance of selecting the right model for the specific problem. This research has the potential to contribute to the early detection and management of CKD, ultimately improving the quality of healthcare for affected individuals.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Aljaaf A. J. (2018). Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics, 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, pp. 1-9, Doi: 10.1109/CEC.2018.8477876.

[2] Aruna O. and Sameerunnisa S., (2023). Chronic Kidney Disease Prediction using Data Pre-Processing Techniques, 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1665-1670, doi: 10.1109/ICSSIT55814.2023.10061025.

[3] Alshakrani S., Taha R. and Hewahi N., (2021). Chronic Kidney Disease Classification Using Machine Learning Classifiers, 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain, 2021, 516-519, doi: 10.1109/3ICT53449.2021.9581345.

[4] Ahmed, A. H., Ahmad, S., Sayed, M. A., Ayon, E. H., Mia, T., & Koli, T. (2023). Predicting the Possibility of Student Admission into Graduate Admission by Regression Model: A Statistical Analysis. *Journal of Mathematics and Statistics Studies*, *4*(4), 97-105.

[5] Aljaaf A. J. et al., (2018) Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics, 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 2018, 1-9, doi: 10.1109/CEC.2018.8477876.

[6] Arif-Ul-Islam and Ripon S. H., (2019). Rule Induction and Prediction of Chronic Kidney Disease Using Boosting Classifiers, Ant-Miner and J48 Decision Tree, 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, 1-6, doi: 10.1109/ECACE.2019.8679388.

[7] Arafat F., Khan T., Bapon A. D., Khan M. I. and Noori S. R. H., (2021). A Deep Learning Approach to Predict Chronic Kidney Disease in Human, 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2021, pp. 1010-1015, doi: 10.1109/IEMCON53756.2021.9623101.

[8] Abu Sa, Tayaba, M., Islam, M. T., Eyasin Ul I P, Tuhin M, Eftekhar H A, Nur N, & Bishnu P G. (2023). Parkinson's Disease Detection through Vocal Biomarkers and Advanced Machine Learning Algorithms. *Journal of Computer Science and Technology Studies*, *5*(4), 142–149. https://doi.org/10.32996/jcsts.2023.5.4.14

[9] Cao, D. M., Amin, M. S., Islam, M. T., Ahmad, S., Haque, M. S., Sayed, M. A., ... & Koli, T. (2023). Deep Learning-Based COVID-19 Detection from Chest X-ray Images: A Comparative Study. *Journal of Computer Science and Technology Studies*, *5*(4), 132-141.

[10] Chittora P. et al., (2021). Prediction of Chronic Kidney Disease - A Machine Learning Perspective, in IEEE Access, vol. 9, pp.17312 17334,2021, Doi: 10.1109/ACCESS.2021.3053763

[11] Desai M. (2019). Early Detection and Prevention of Chronic Kidney Disease, 2019 5th International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, India, 2019, pp. 1-5, Doi: 10.1109/ICCUBEA47591.2019.9128424.

[12] Dissanayake D. V., Sobana S., Yogarajah B. and Nagulan R., (2023). Chronic Kidney Disease Detection using Machine Learning Algorithms: A Sri Lankan Study, 2023 3rd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2023, 60-65, doi: 10.1109/ICARC57651.2023.10145656.

[13] Emon M. U., Imran A. M., Islam R., Keya M. S., Zannat R. and Ohidujjaman, (2021). Performance Analysis of Chronic Kidney Disease through Machine Learning Approaches, 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, 713-719, doi: 10.1109/ICICT50816.2021.9358491.

[14] Emon M. U., Imran A. M., Islam R., Keya M. S., Zannat R. and Ohidujjaman, (2021). Performance Analysis of Chronic Kidney Disease through Machine Learning Approaches, 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, 713-719, Doi: 10.1109/ICICT50816.2021.9358491.

[15] Ekanayake I. U. and Herath D., (2020). Chronic Kidney Disease Prediction Using Machine Learning Methods, 2020 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 2020, pp. 260-265, Doi: 10.1109/MERCon50084.2020.9185249.

[16] Ghosh P., Javed M S F. M., Shultana S., Afrin S., Anjum A. A. and Khan A. A. (2020). Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm, 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Bangkok, Thailand, 2020, pp. 1-6, Doi: 10.1109/iSAI-NLP51646.2020.9376787.

[17] Haque, M. S., Amin, M. S., Miah, J., Cao, D. M., & Ahmed, A. H. (2023). Boosting Stock Price Prediction with Anticipated Macro Policy Changes. *Journal of Mathematics and Statistics Studies*, *4*(3), 29–34. https://doi.org/10.32996/jmss.2023.4.3.4

[18] Islam M. A., Akter S., Hossen S., Keya S. A., Tisha S. A. and Hossain S., (2020). Risk Factor Prediction of Chronic Kidney Disease based on Machine Learning Algorithms, 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 952-957, Doi: 10.1109/ICISS49785.2020.9315878.

[19] Islam, M, N, Shah A A., Islam, M., Rakib R, Abdur M., Miah J., Kayyum S., Shadaab, A ., Faisal, F al et al. (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. International Conference on Machine Learning, Image Processing, Network Security and Data Sciences. Springer, Singapore, 2020.

[20] Jhumka K., Auzine M. M., Casseem M. S., Khan M. H. -M. and Mungloo-Dilmohamud Z. (2022). Chronic Kidney Disease Prediction using Deep Neural Network, 2022 3rd International Conference on Next Generation Computing Applications (NextComp), Flic-en-Flac, Mauritius, 2022, pp. 1-5, Doi: 0.1109/NextComp55567.2022.9932200.

[21] Jonayet M, Razib H K, Sabbir A, and Isatyaq M, (2023). A comparative study of detecting covid 19 by using chest X-ray images – A deep learning approach, In 2023 IEEE World AIIoT Congress (AIIoT) Regular Research Paper, 2023.

[22] Khan R. H. and Miah J., (2022). Performance Evaluation of a new one-time password (OTP) scheme using stochastic petri net (SPN), 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2022, 407-412, doi: 10.1109/AIIoT54504.2022.9817203.

[23] Khan R. H., Miah J., Rahman M. M., Hasan M. M. and Mamun M. (2023). A study of forecasting stocks price by using deep Reinforcement Learning, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, 0250-0255, Doi: 10.1109/AIIoT58121.2023.10174358.

[24] Khan R. H., Miah J., Abed N S. A. and Islam M. (2023). A Comparative Study of Machine Learning Classifiers to Analyze the Precision of Myocardial Infarction Prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, pp. 0949-0954, Doi: 10.1109/CCWC57344.2023.10099059.

[25] Khan R.H., Miah R.H., Tayaba M., Rahman M.M (2023). A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), USA.

[26] Kayyum S. et al., (2020). Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-7, Doi: 10.1109/ICCCI48352.2020.9104104.

[27] Kashyap C. P., Dayakar R G. S. and Balamurugan M., (2022). Prediction of Chronic Disease in Kidneys Using Machine Learning Classifiers, 2022 1st International Conference on Computational Science and Technology (ICCST), CHENNAI, India, 2022, 562-567, doi: 10.1109/ICCST55948.2022.10040329.

[28] Khan R. H., Miah J., Arafat S. M. Y., Syeed M. M. M. and Ca D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks, 2023 15th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2023, 104-109, doi: 10.1109/IIT59782.2023.10366426.

[29] Miah, J., Haque, M. S., Cao, D. M., & Sayed, M. A. (2023). Enhancing Traffic Density Detection and Synthesis through Topological Attributes and Generative Methods. *Journal of Computer Science and Technology Studies*, *5*(4), 69–77. https://doi.org/10.32996/jcsts.2023.5.4.8

[30] Mia, M. T., Ray, R. K., Ghosh, B. P., Chowdhury, M. S., Al-Imran, M., Das, R., Sarkar, M., Sultana, N., Nahian, S. A., & Puja, A. R. (2023). Dominance of External Features in Stock Price Prediction in a Predictable Macroeconomic Environment. *Journal of Business and Management Studies*, *5*(6), 128–133. https://doi.org/10.32996/jbms.2023.5.6.10

[31] Miah J., Ca D. M., Sayed M. A., Lipu E. R., Mahmud F. and Arafat S. M. Y., (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction, 2023 15th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2023, 49-54, doi: 10.1109/IIT59782.2023.10366476.

[32] Miah J., Mamun M., Rahman M.M., Mahmud M. I., Islam A. M., Ahmad S., (2022). MHfit: Mobile Health Data for Predicting Athletics Fitness using Machine Learning Models 2022 2nd International seminar on machine learning, Optimization, and Data Science (ISMODE), 2022, (Preprint)

[33] Nikhila, (2021). Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm, 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCIS), Greater Noida, India, 2021, pp. 476-480, Doi: 10.1109/ICCCIS51004.2021.9397144.

[34] Nandhini G. and Aravinth J., (2021). chronic kidney disease prediction using machine learning techniques, 2021 International Conference on Recent Trends on Electronics, Information, Communication &Technology (RTEICT), Bangalore, India, 2021, pp. 227–232, Doi: 10.1109/RTEICT52294.2021.9573971.

[35] Rajeshwari and Yogish H. K. (2022). Prediction of Chronic Kidney Disease Using Machine Learning Technique, 2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP), Bengaluru, India, 2022, 1-6, doi: 10.1109/CCIP57447.2022.10058678.

[36] Rahman T. M., Siddiqua S., Rabby S. E., Hasan N. and Imam M. H. (2019). Early Detection of Kidney Disease Using ECG Signals Through Machine Learning Based Modelling, 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 2019, pp. 319-323, Doi: 10.1109/ICREST.2019.8644354.

[37] Salekin A. and Stankovic J., (2016). Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes, 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL, USA, 2016, pp. 262-270, Doi: 10.1109/ICHI.2016.36.

[38] Shukla G., Dhuriya G., Pillai S. K. and Saini A., (2023). Chronic Kidney Disease Prediction Using Machine Learning Algorithms and the Important Attributes for the Detection, 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom, 2023. 1-4, doi: 10.1109/GlobConET56651.2023.10149900.

[39] Snegha J., Tharani V., Preetha S. D., Charanya R. and Bhavani S. (2020). Chronic Kidney Disease Prediction Using Data Mining, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 2020, 1-5, doi: 10.1109/ic-ETITE47903.2020.482.

[40] Sarkar, M., Ayon, E. H., Mia, M. T., Ray, R. K., Chowdhury, M. S., Ghosh, B. P., Al-Imran, M., Islam, M. T., Tayaba, M., & Puja, A. R. (2023). Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases. *Journal of Computer Science and Technology Studies*, *5*(4), 186–193. https://doi.org/10.32996/jcsts.2023.5.4.19

[41] Tayaba, M., Ayon, E. H., Mia, M. T., Sarkar, M., Ray, R. K., Chowdhury, M. S., Al-Imran, M., Nobe, N., Ghosh, B. P., Islam, M. T., & Puja, A. R. (2023). Transforming Customer Experience in the Airline Industry: A Comprehensive Analysis of Twitter Sentiments Using Machine Learning and Association Rule Mining. *Journal of Computer Science and Technology Studies*, *5*(4), 194–202. https://doi.org/10.32996/jcsts.2023.5.4.20

[42] Yildirim P., (2017). Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction, 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 2017, pp. 193-198, doi: 10.1109/COMPSAC.2017.84.