

# Predictive Modeling Of Chronic Kidney Disease Using Machine Learning Techniques

Dr.Gokila Vani , S.Surya Kalyan, K.Ajay Kumar, K.Sujith Kumar

<sup>1</sup>Assistant Professor, Department Of IT, Guru Nanak Institutions Technical Campus (Autonomous), India.

<sup>2,3,4</sup>B.Tech Students, Department Of IT, Guru Nanak Institutions Technical Campus (Autonomous), India

## ABSTRACT

*Chronic Kidney Disease is one of the most critical illness nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. With the help of a machine learning classifier algorithms, the doctor can detect the disease on time. For this perspective, Chronic Kidney Disease prediction has been discussed in this article. Chronic Kidney Disease dataset has been taken from the UCI repository. Seven classifier algorithms have been applied in this research such as artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 & with penalty L2 and random tree. The important feature selection technique was also applied to the dataset. For each classifier, the results have been computed based on (i) full features, (ii) correlation-based feature selection, (iii) Wrapper method feature selection, (iv) Least absolute shrinkage and selection operator regression, (v) synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features, (vi) synthetic minority oversampling technique with full features. From the results, it is marked that LSVM with penalty L2 is giving the highest accuracy of 98.86% in synthetic minority over-sampling technique with full features. Along with accuracy, precision, recall, F-measure, area under the curve and GINI coefficient have been computed and compared results of various algorithms have been shown in the graph. Least absolute shrinkage and selection operator regression selected features with synthetic minority over-sampling technique gave the best after synthetic minority over-sampling technique with full features. In the synthetic minority over-sampling technique with least absolute shrinkage and selection operator selected features, again linear support vector machine gave the highest accuracy of 98.46%. Along with machine learning models one deep neural network has been applied on the same dataset and it has been noted that deep neural network achieved the highest accuracy of 99.6%.*

## 1-INTRODUCTION

Chronic kidney Disease (CKD) means your kidneys are damaged and not filtering your blood the way it should. The primary role of kidneys is to filter extra

water and waste from your blood to produce urine and if the person has suffered from CKD, it means that wastes are collected in the body. This disease is chronic because of the damage gradually over a long period. It is flatterer a common disease worldwide [1]. Due to CKD may have some health troubles. There are many causes for CKD like diabetes, high blood pressure, heart disease. Along with these critical diseases, CKD also depends on age and gender. If your kidney is not working, then you may notice one or more symptoms like abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, vomiting. There are two main diseases of CKD: (i) diabetes and (ii) high blood pressure. So that controlling of these two diseases is the prevention of CKD. Usually, CKD does not give any sign till kidney is damaged badly. CKD is being increased rapidly as per the studies hospitalization cases increase 6.23 per cent per year but the global mortality rate remains fixed [4]. There are few diagnostic tests to check the condition of CKD: (i) estimated glomerular filtration rate(eGFR) (ii) urine test (iii) blood pressure. The doctor also asks for urine test for kidney functionality because kidneys make urine. If the urine contains blood and protein [6], that means your kidney is not working properly. Doctor measures blood pressure as Blood pressure range shows how your heart is pumping blood. If eGFR value reaches less than 15, that means the patient has end-stage kidney disease. At this point, there are only available treatments: (i) dialysis and (ii) kidney transplant. Patient's life after dialysis depends on such factors as age, gender, frequency and duration of dialysis, physical movement of the body and mental health [7]. If dialysis is not possible, the doctor has only one solution, i.e., kidney transplantation. However, it is extremely expensive [8]. Therefore, it is critical noteworthy in early recognition, monitoring and handling of the disease. It is essential to predict the striding of CKD with appropriate accuracy due to its dynamic and secretive nature in the early stages and patient abnormality. Medical treatment of CKD is prescribed by the stage. Anything other than this, it is very imperative to characterize the organization of the infection because it gives a few indications. It underpins the assurance of fundamental intercessions and medications. Medical treatment is a very significant application area of intellectual intelligent systems [10]. Afterwards, Data mining can play a big role to find out hidden information

from the huge patient medical and treatment dataset that doctors frequently obtain from patients to get pieces of knowledge about the symptomatic data and to execute precise treatment plans. Data mining can be categorized as the method of extracting hidden information from a huge dataset. Data mining strategies are connected and utilized broadly in various contexts and areas. Using data mining methods, we may predict, classify, filter and cluster data. The objective states the algorithm processing of a training set containing a set of attributes and targets. Data mining is suitable to mining in data if the dataset is huge but we can also do it with the help of machine learning with a small dataset. The machine learning can also find data analysis and pattern detection [9]. A variety of health dataset is present so machine learning algorithms are best fit to improve the accuracy of diagnosis prediction [11]. As healthcare electronic dataset grows rapidly, machine learning algorithms are becoming more common in healthcare. Qin et al. proposed data assertion and sample diagnosis achievable in CKD diagnosis. KNN is used for data assertion. Six classifiers algorithms used for accuracy of diagnosis: logistic regression, random forest, support vector machine, K-nearest neighbor, naive Bayes classifier and feed-forward neural network. In these classifiers random forest gives better accuracy, i.e., 99.75%. Vasquez-Morales et al. [14] developed a neural network model for risk prediction of chronic kidney disease development on 40000 instances dataset and their model accuracy was 95%.

## 2-LITERATURE SURVEY

**Title:** Predictive Analytics Of Chronic Kidney Disease By Using Machine Learning

**AUTHORS:** Khan, P. F., Reddy, M. R., Samatha, K., Chowdary, R. A., & Rao, P. P.

**Year:** 2020

Predictive analytics for healthcare using machine learning is a challenged task to help doctors decide the exact treatments for saving lives. In this paper, we present machine learning techniques for predicting the chronic kidney disease using clinical data. Four machine learning methods are explored including K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers. These predictive models are constructed from chronic kidney disease dataset and the performance of these models are compared together in order to select the best classifier for predicting the chronic kidney disease.

**Title:** Comparative Analysis of Machine Learning Methods to Detect Chronic Kidney Disease

**AUTHORS:** Roy, M. S., Ghosh, R., Goswami, D., & Karthik, R.

**Year:** 2020

Chronic kidney disease (CKD) is a major burden on the healthcare system because of its increasing prevalence, high risk of progression to end-stage renal disease, and poor morbidity and mortality prognosis. It is rapidly becoming a global health crisis. Unhealthy dietary habits and insufficient water consumption are significant contributors to this disease. Without kidneys, a person can only live for 18 days on average, requiring kidney transplantation and dialysis. It is critical to have reliable techniques at predicting CKD in its early stages. Machine learning (ML) techniques are excellent in predicting CKD. The current study offers a methodology for predicting CKD status using clinical data, which incorporates data preprocessing, a technique for managing missing values, data aggregation, and feature extraction. A number of physiological variables, as well as ML techniques such as logistic regression (LR), decision tree (DT) classification, and -nearest neighbor (KNN), were used in this work to train three distinct models for reliable prediction. The LR classification method was found to be the most accurate in this role, with an accuracy of about 97 percent in this study. The dataset that was used in the creation of the technique was the CKD dataset, which was made available to the public. Compared to prior research, the accuracy rate of the models employed in this study is considerably greater, implying that they are more trustworthy than the models used in previous studies as well. A large number of model comparisons have shown their resilience, and the scheme may be inferred from the study's results.

**Title:** Prediction of chronic kidney disease-a machine learning perspective

**AUTHORS:** Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., & Bolshev, V.

**Year:** 2020

Chronic Kidney Disease is one of the most critical illness nowadays and proper diagnosis is required as soon as possible. Machine learning technique has become reliable for medical treatment. With the help of a machine learning classifier algorithms, the doctor can detect the disease on time. For this perspective, Chronic Kidney Disease prediction has been discussed in this article. Chronic Kidney Disease dataset has been taken from the UCI repository. Seven classifier algorithms have been applied in this research such as artificial neural network, C5.0, Chi-square Automatic interaction detector, logistic regression, linear support vector machine with penalty L1 & with penalty L2 and random tree. The important feature selection technique was also applied to the dataset. For each classifier, the results have been computed based on (i) full features, (ii) correlation-based feature

selection, (iii) Wrapper method feature selection, (iv) Least absolute shrinkage and selection operator regression, (v) synthetic minority over-sampling technique with least absolute shrinkage and selection operator regression selected features, (vi) synthetic minority over-sampling technique with full features. From the results, it is marked that LSVM with penalty L2 is giving the highest accuracy of 98.86% in synthetic minority over-sampling technique with full features. Along with accuracy, precision, recall, F-measure, area under the curve and GINI coefficient have been computed and compared results of various algorithms have been shown in the graph. Least absolute shrinkage and selection operator regression selected features with synthetic minority over-sampling technique gave the best after synthetic minority over-sampling technique with full features. In the synthetic minority over-sampling technique with least absolute shrinkage and selection operator selected features, again linear support vector machine gave the highest accuracy of 98.46%.

**Title:** Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques

**AUTHORS:** Janani, J., & Sathyaraj, R

**Year:** 2020

The chronic kidney failure is a serious health issue and if not detected and treated at the early stages, it can be very deadly. Hence the major objective of this paper is to develop a reliable machine learning model which predicts the CKD with a high accuracy rate. The CKD data set is downloaded from the famous UCI ML repository but it suffers from a lot of missing values. To handle the missing values KNN Imputation is used. Feature selection is also performed with the help of information gain as the dataset is huge and hence the cost of modelling can be very high. Various other pre-processing steps like label encoding and Min-max normalization is performed to attain a clean dataset. After pre-processing, various ML algorithms like logistic regression, naïve bayes, artificial neural network and random forest are applied and their performances are compared with the help of various performance metrics. A hybrid of Random Forest and Adaboost algorithm is proposed and it achieves a better accuracy when compared to the other individual component models and hence it can be proved that the proposed hybrid model is much better and accurate in diagnosing CKD.

**Title:** Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm

**AUTHORS:** Wang, W., Chakraborty, G., & Chakraborty, B

**Year:** 2020

Creatinine is a type of metabolite of blood that is strongly correlated to glomerular filtration rate

(GFR). As measuring GFR is difficult, creatinine value is used for indirectly determining GFR and then the stage of chronic kidney disease (CKD). Adding a creatinine test into routine health examination could detect CKD. As more items for comprehensive examination means higher cost, creatinine testing is not included in the routine health examination in many countries. An algorithm based on common test results, without creatinine test, to evaluate the risk of CKD will increase the chance of its early detection and treatment. Methods: In this study, we used open source data containing 1 million samples. These data contain 23 health-related features, including common diagnostic test results provided by National Health Insurance Sharing Service (NHSS). A low GFR indicates possible chronic kidney disease (CKD). As is commonly accepted in the medical community, a GFR of 60 mL/min is used as the threshold, below which is considered to have CKD. In this study, the first step aims to build a regression model to predict the value of creatinine from 23 features, and then combine the predicted value of creatinine with the original 23 features to evaluate the risk of CKD. We will show by simulation that by the proposed method we can achieve better prediction results compared to direct prediction from 23 features. The data is extremely unbalanced for predicting the target variable creatinine. We used undersampling method and proposed a new cost-sensitive mean-squared error (MSE) loss function to deal with the problem. Regrading model selection, this work used three machine learning models: a bagging tree model named Random Forest, a boosting tree model named XGBoost, and a neural network based model named ResNet. To improve the result of the creatinine predictor, we averaged results from eight predictors, a method known as ensemble learning. Finally, the predicted creatinine and the original 23 features is used to predict the risk of CKD. Results: We optimized results of R-Squared (R<sup>2</sup>) value to select the appropriate undersampling strategy and the regression model for the regression stage of creatinine prediction. Ensembled model achieved the best performance of R<sup>2</sup> of 0.5590. The six factors from 23 are selected from the top of the list of how strongly they affect the creatinine value. They are sex, age, hemoglobin, the level of urine protein, waist circumference, and habit of smoking. Using the predicted value of creatinine, an area under Receiver Operating Characteristic curve (AUC) of 0.76 is achieved while classifying samples for CKD. Conclusions: Using commonly available health parameters, the proposed system can assess the risk of CKD for public health. High-risk subjects can be screened and advised to take a creatinine test for further confirmation. In this way, we can reduce

the impact of CKD on public health and facilitate early detection for many, where a blanket test of creatinine is not available for all.

### 3-METHODOLOGY

Chronic Kidney Disease (CKD) is a non-communicable illness that affects 10-15% of the global population and significantly impacts morbidity, mortality, and hospital admission rates. The kidneys play a vital role in excretion and osmoregulation, removing harmful and excess substances from the body. CKD, characterized by a long-term decline in kidney function, can lead to severe complications such as end-stage renal disease and even death if not diagnosed and treated early. Early and accurate detection of CKD stages is essential to reduce its adverse effects on patients' health. This project aims to develop a system for predicting CKD using machine learning techniques, specifically an Artificial Neural Network (ANN). The system is trained and tested on the chronic kidney disease dataset from the UCI Machine Learning Repository. To make the system more accessible, a web application is built using the Flask framework, allowing users to input their details and predict the presence of CKD. By utilizing machine learning for medical diagnosis, this project contributes to improving early CKD prediction and demonstrates the potential of such techniques in enhancing healthcare outcomes.

#### 1. Data Transformation into Knowledge Graphs:

- Data Collection and Preprocessing: Begin with the collection of comprehensive employee data, including demographic details, job roles, performance metrics, and engagement scores. Preprocess this data to ensure consistency, handle missing values, and normalize numerical features.
- Graph Construction: Transform the pre-processed tabular data into a knowledge graph by mapping employees, departments, job roles, and other relevant entities as nodes, and defining relationships (e.g., reporting structures, project collaborations) as edges. This graph structure captures the complex interrelations within the organization, providing a richer dataset for analysis.

#### 2. Hybrid Predictive Modelling:

- Enhanced Linear Support Vector Machine (L-SVM): Integrate the features extracted by the GCN into an enhanced L-SVM model. This hybrid approach leverages the linear separability of SVMs, augmented by the deep relational insights provided by the GCN, to improve turnover prediction accuracy.

- Benchmarking Against Traditional Models: Conduct a comparative analysis by training and testing the hybrid model against traditional machine learning algorithms such as Random Forest, Logistic Regression, and Gradient Boosting Machines. This benchmarking assesses the performance gains achieved through the integration of knowledge graphs and GCNs.

#### 3. Explainable AI (XAI) Integration:

- Feature Importance Analysis: Implement XAI techniques, such as SHAP (Shapley Additive explanations) values, to interpret the contribution of each feature to the model's predictions. This analysis identifies the key factors influencing turnover, offering transparency and actionable insights.
- Visual Explanation Tools: Develop visual dashboards that illustrate the relationships between key factors and turnover risk, enabling HR professionals to easily understand and communicate the reasons behind model predictions.

#### 4. Empirical Evaluation:

- Dataset Analysis: Apply the methodology to a real-world dataset of 1,470 IBM employees. Perform data splitting into training and testing sets, ensuring robust evaluation metrics such as accuracy, precision, recall, and F1-score.
- Case Study and Scenario Testing: Conduct case studies to validate the model's performance in different organizational scenarios, such as changes in management structure or large-scale onboarding, to assess the model's adaptability and reliability.

#### 5. Continuous Learning and Adaptation:

- Model Retraining and Update: Implement a system for continuous learning where the model is periodically retrained with new data to adapt to evolving organizational dynamics.

## 4-REQUIREMENTS ENGINEERING

### HARDWARE REQUIREMENTS

The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It should state what the system do and not how it should be implemented.

### HARDWARE

- PROCESSOR : PENTIUM IV  
2.6 GHz, Intel Core 2 Duo.
- RAM : 512 MB DD  
RAM
- MONITOR : 15" COLOR

- HARD DISK : 40 GB

### SOFTWARE REQUIREMENTS

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

- Front End : J2EE  
(JSP, SERVLET)
- Back End : MY  
SQL 5.5
- Operating System :  
Windows 10
- IDE :  
Eclipse

### FUNCTIONAL REQUIREMENTS

A functional requirement defines a function of a software-system or its component. A function is described as a set of inputs, the behavior, Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture.

### NON-FUNCTIONAL REQUIREMENTS

**The major non-functional Requirements of the system are as follows**

### CLASS DIAGRAM

### Usability

The system is designed with completely automated process hence there is no or less user intervention.

### Reliability

The system is more reliable because of the qualities that are inherited from the chosen platform python. The code built by using python is more reliable.

### Performance

This system is developing in the high level languages and using the advanced back-end technologies it will give response to the end user on client system with in very less time.

### Supportability

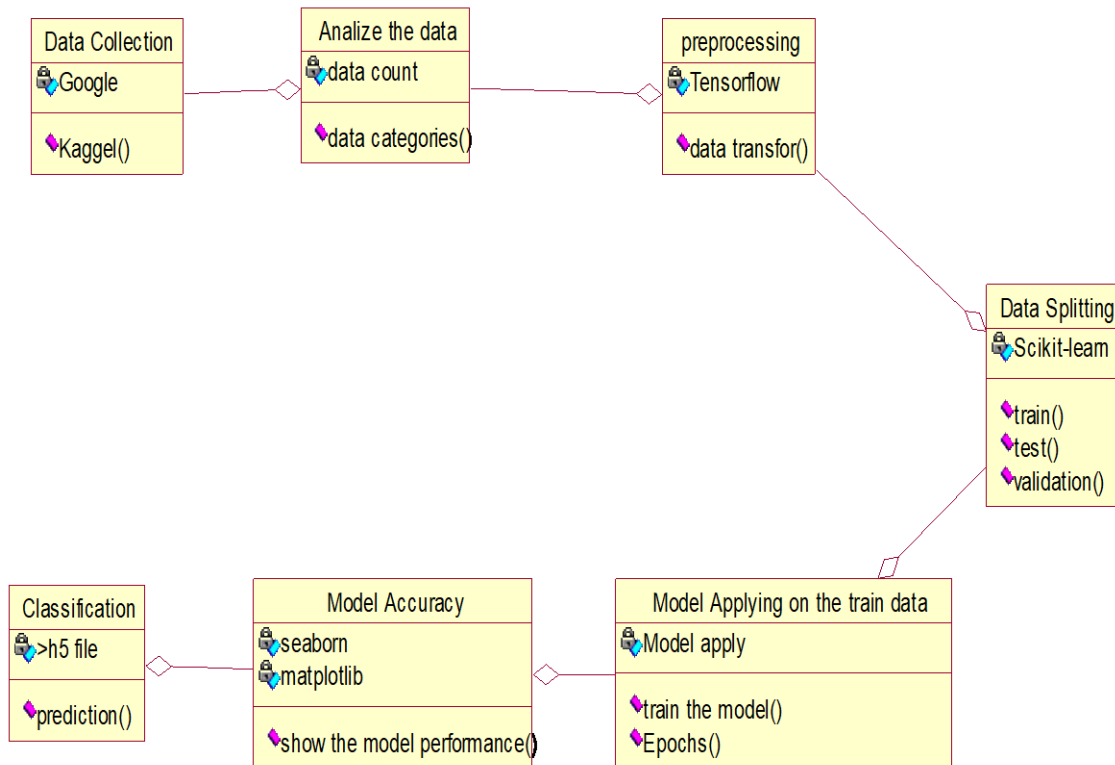
The system is designed to be the cross platform supportable. The system is supported on a wide range of hardware and any software platform, which is built into the system.

### Implementation

The system is implemented in web environment using Jupyter notebook software. The server is used as the intelligence server and windows 10 professional is used as the platform. Interface the user interface is based on Jupyter notebook provides server system.

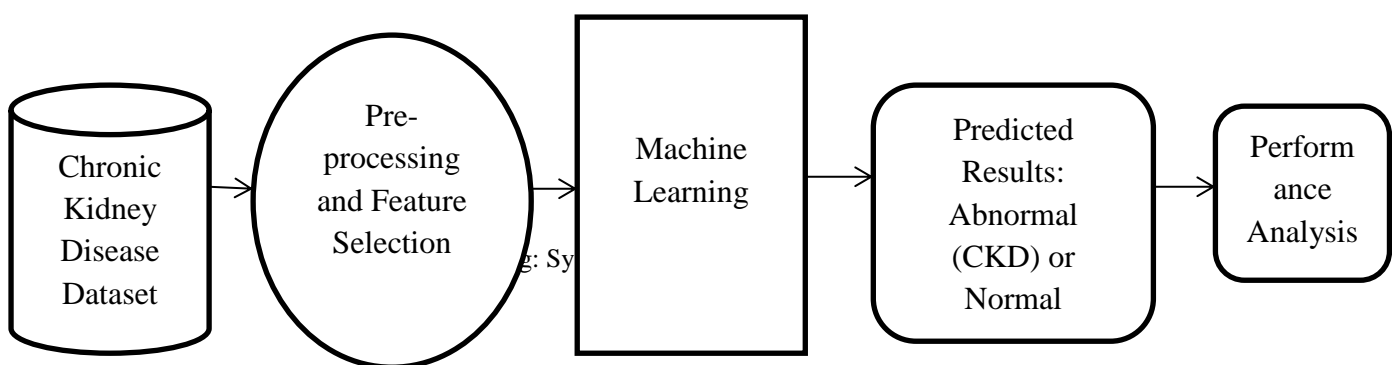
### 5-DESIGN ENGINEERING

Design Engineering deals with the various UML [Unified Modelling language] diagrams for the implementation of project. Design is a meaningful engineering representation of a thing that is to be built. Software design is a process through which the requirements are translated into representation of the software. Design is the place where quality is rendered in software engineering.



In this class diagram represents how the classes with attributes and methods are linked together to perform the verification with security. From the above diagram shown the various classes involved in our project.

**SYSTEM ARCHITECTURE:**



**6-IMPLEMENTATION**

**LIST:**

Lists are implemented in the JCF via the java.util.List interface. It defines a list as essentially a more flexible version of an array. Elements have a specific order, and duplicate elements are allowed. Elements can be placed in a specific position. They can also be searched for

within the list. Two concrete classes implement List. The first is java.util.ArrayList, which implements the list as an array. Whenever functions specific to a list are required, the class moves the elements around within the array in order to do it. The other implementation is java.util.LinkedList. This class stores the elements in nodes that each have a pointer to the previous and next nodes in the list. The list can

be traversed by following the pointers, and elements can be added or removed simply by changing the pointers around to place the node in its proper place.

#### SET:

Java's [java.util.Set](#) interface defines the set. A set can't have any duplicate elements in it. Additionally, the set has no set order. As such, elements can't be found by index. Set is implemented by [java.util.HashSet](#), [java.util.LinkedHashSet](#), and [java.util.TreeSet](#). HashSet uses a hash table. More specifically, it uses a [java.util.HashMap](#) to store the hashes and elements and to prevent duplicates. [java.util.LinkedHashSet](#) extends this by creating a doubly linked list that links all of the elements by their insertion order. This ensures that the iteration order over the set is predictable. [java.util.TreeSet](#) uses a red-black tree implemented by a [java.util.TreeMap](#). The red-black tree makes sure that there are no duplicates. Additionally, it allows TreeSet to implement [java.util.SortedSet](#).

The [java.util.Set](#) interface is extended by the [java.util.SortedSet](#) interface. Unlike a regular set, the elements in a sorted set are sorted, either by the element's `compareTo()` method, or a method provided to the constructor of the sorted set. The first and last elements of the sorted set can be retrieved, and subsets can be created via minimum and maximum values, as well as beginning or ending at the beginning or ending of the sorted set. The [SortedSet](#) interface is implemented by [java.util.TreeSet](#). [java.util.SortedSet](#) is extended further via the [java.util.NavigableSet](#) interface. It's similar to [SortedSet](#), but there are a few additional methods. The `floor()`, `ceiling()`, `lower()`, and `higher()` methods find an element in the set that's close to the parameter. Additionally, a descending iterator over the items in the set is provided. As with [SortedSet](#), [java.util.TreeSet](#) implements [NavigableSet](#).

#### MAP:

Maps are defined by the [java.util.Map](#) interface in Java. Maps are simple data structures that associate a key with a value. The element is the value. This lets the map be very flexible. If the key is the hash code of the element, the map is essentially a set. If it's just an increasing number, it becomes a list. Maps are implemented by [java.util.HashMap](#), [java.util.LinkedHashMap](#), and [java.util.TreeMap](#). [HashMap](#) uses a hash table. The hashes of the keys are used to find the values in various buckets. [LinkedHashMap](#) extends this by creating a doubly linked list between the elements. This allows the elements to be accessed in the order in which they were inserted into the map. [TreeMap](#), in contrast to [HashMap](#) and [LinkedHashMap](#), uses a red-black tree. The keys are used as the values for

the nodes in the tree, and the nodes point to the values in the map

#### THREAD:

Simply put, a *thread* is a program's path of execution. Most programs written today run as a single thread, causing problems when multiple events or actions need to occur at the same time. Let's say, for example, a program is not capable of drawing pictures while reading keystrokes. The program must give its full attention to the keyboard input lacking the ability to handle more than one event at a time. The ideal solution to this problem is the seamless execution of two or more sections of a program at the same time.

### 7-CONCLUSION

This article objects to predict Chronic Kidney Disease based on full features and important features of CKD dataset. For feature selection three different techniques have been applied: correlation-based feature selection, Wrapper method and LASSO regression. In this perception, seven classifiers algorithm were applied viz. artificial neural network, C5.0, logistic regression, CHAID, linear support vector machine (LSVM), K-Nearest neighbors and random tree. For each classifier, the results were computed based on full features, selected features by CFS, selected features by Wrapper, selected features by LASSO regression, SMOTE with selected features by LASSO, SMOTE with full features. It was observed that LSVM achieved the highest accuracy of 98.86% in SMOTE with full features. All classifiers algorithms performed well on features selected by LASSO regression with SMOTE and without SMOTE. SMOTE with full features gave the best result for all 5 classifiers. In this research, a total of 7 classifiers were used. However, Logistic and KNN did not give suitable results and it was why they were not used in SMOTE. As per the result, it is concluded that SMOTE is a best technique for balancing a dataset. It is noted that SMOTE gave better results with selected features by LASSO regression as compare to without SMOTE on LASSO regression model. LSVM achieved the highest accuracy in all experiments as compared to other classifiers algorithms.

### REFERENCES

- [1] Q.-L. Zhang and D. Rothenbacher, "Prevalence of chronic kidney disease in population-based studies: Systematic review," *BMC Public Health*, vol. 8, no. 1, p. 117, Dec. 2008.
- [2] W. M. McClellan, D. G. Warnock, S. Judd, P. Muntner, R. Kewalramani, M. Cushman, L. A. McClure, B. B. Newsome, and G. Howard, "Albuminuria and racial disparities in the risk for

- ESRD,” *J. Amer. Soc. Nephrol.*, vol. 22, no. 9, pp. 1721–1728, Aug. 2011.
- [3] M. K. Haroun, “Risk factors for chronic kidney disease: A prospective study of 23,534 men and women in Washington County, Maryland,” *J. Amer. Soc. Nephrol.*, vol. 14, no. 11, pp. 2934–2941, Nov. 2003.
- [4] W. D. Souza, L. C. D. Abreu, L. G. D. Silva, and I. M. P. Bezerra, “Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017,” *Wisit Cheungpasitporn, Univ. Mississippi Medical Center, Rochester, MN, USA, Tech. Rep.*, 2019, doi: 10.1371/journal.pone.0224889.
- [5] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, “Estimated glomerular filtration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice,” *Oman Med. J.*, vol. 27, no. 4, pp. 339–340, 2012.
- [6] A. S. Levey, D. Catran, A. Friedman, W. G. Miller, J. Sedor, K. Tuttle, B. Kasiske, and T. Hostetter, “Proteinuria as a surrogate outcome in CKD: Report of a scientific workshop sponsored by the national kidney foundation and the US food and drug administration,” *Amer. J. Kidney Diseases*, vol. 54, no. 2, pp. 205–226, Aug. 2009.
- [7] S. Gerogianni, “Concerns of patients on dialysis: A research study,” *Health Sci. J.*, vol. 8, no. 4, pp. 423–437, 2014.
- [8] J. R. Chapman, “What are the key challenges we face in kidney transplantation today?” *Transplantation Res.*, vol. 2, no. S1, pp. 1–7, Nov. 2013.
- [9] T. Xiuyi and G. Yuxia, “Research on application of machine learning in data mining,” in *Proc. IOP Conf., Mater. Sci. Eng.*, 2018, doi: 10.1088/1757-899X/392/6/06220.
- [10] B. Zupan, A. J. Halter, and M. Bohanec, “Qualitative model approach to computer assisted reasoning in physiology,” in *Proc. Intell. Data Anal. Med. Pharmacol. (IDAMAP)*, Brighton, U.K., 2018, pp. 1–7.
- [11] A. Dhillon and A. Singh, “Machine learning in healthcare data analysis: A survey,” *J. Biol. Today’s World*, vol. 8, no. 2, pp. 1–10, Jan. 2018.
- [12] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, and R. Ranganath, “A review of challenges and opportunities in machine learning for health,” in *Proc. AMIA Joint Summits Transl. Sci.*, 2020, p. 191.
- [13] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, “A machine learning methodology for diagnosing chronic kidney disease,” *IEEE Access*, vol. 8, pp. 20991–21002, 2020.
- [14] G. R. Vasquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, “Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning,” *IEEE Access*, vol. 7, pp. 152900–152910, 2019.
- [15] Z. Chen, X. Zhang, and Z. Zhang, “Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models,” *Int. Urol. Nephrol.*, vol. 48, no. 12, pp. 2069–2075, Jun. 2016.
- [16] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, “Analysis of chronic kidney disease dataset by applying machine learning methods,” in *Proc. IEEE 12th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Oct. 2018, pp. 1–4.
- [17] K. R. A. Padmanaban and G. Parthiban, “Applying machine learning techniques for predicting the risk of chronic kidney disease,” *Indian J. Sci. Technol.*, vol. 9, no. 29, Aug. 2016.
- [18] L. Kilvia De Almeida, L. Lessa, A. Peixoto, R. Gomes, and J. Celestino, “Kidney failure detection using machine learning techniques,” in *Proc. 8th Int. Workshop ADVANCES ICT Infrastructures Services*, 2020, pp. 1–8.
- [19] W. Gunarathne, K. D. M Perera, and K. A. D. C. P Kahandawaarachchi, “Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD),” in *Proc. IEEE 17th Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2017, pp. 291–296.
- [20] H. Polat, H. D. Mehr, and A. Cetin, “Huseyin polat1 & homay danaei mehr1 & aydin cetin,” *J. Med. Syst.*, vol. 41, no. 4, pp. 1–11, Apr. 2017.
- [21] S. Drall, G. S. Drall, S. Singh, and B. B. Naib, “Chronic kidney disease prediction using machine learning: A new approach,” *Int. J. Manage., Technol. Eng.*, vol. 8, pp. 278–287, May 2018.
- [22] M. Almasoud and T. E. Ward, “Detection of chronic kidney disease using machine learning algorithms with least number of predictors,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 8, pp. 89–96, 2019.
- [23] S. Shankar, S. Verma, S. Elavarthy, T. Kiran, and P. Ghuli, “Analysis and prediction of chronic kidney disease,” *Int. Res. J. Eng. Technol.*, vol. 7, no. 5, May 2020, pp. 4536–4541.
- [24] S. Vijayarani1 and S. Dhayanand, “Kidney disease prediction using SVM and ANN algorithms,” *Int. J. Comput. Bus. Res.*, vol. 6, no. 2, pp. 1–12, Mar. 2015.
- [25] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, and Z. Ye, “Comparison and development of machine learning tools in the prediction of chronic kidney disease progression,” *J. Transl. Med.*, vol. 17, p. 119, Dec. 2019.
- [26] M. S. Gharibdousti, K. Azimi, S. Hathikal, and D. H. Won, “Prediction of chronic kidney disease using data mining techniques,” in *Proc. Ind. Syst.*

Eng. Conf., K. Coperich, E. Cudney, H. Nembhard, Eds., 2017, pp. 2135–2140.

[27] E. M. Karabulut, S. A. Ozel, and T. Ibriki, “A comparative study on the effect of feature selection on classification accuracy,” *Procedia Technol.*, vol. 1, pp. 323–327, Jan. 2012.

[28] A. Wosiak and D. Zakrzewska, “Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis,” *Complexity*, vol. 2018, Oct. 2018, Art. no. 2520706.

[29] N. A. Nnamoko, F. N. Arshad, D. England, J. Vora, and J. Norman, “Evaluation of filter and wrapper methods for feature selection in supervised machine learning,” in *Proc. 15th Annu. Postgraduate Symp. Converg. Telecommun., Netw. Broadcast.*, Liverpool, U.K., 2014, pp. 2–33.

[30] J. M. Pereira, M. Basto, and A. F. D. Silva, “The logistic lasso and ridge regression in predicting corporate failure,” *Procedia Econ. Finance*, vol. 39, pp. 634–641, Jan. 2016.

[31] P. G. Scholar, “Chronic kidney disease prediction using machine learning,” *Int. J. Eng. Res. Technol.*, vol. 9, no. 7, pp. 137–140, 2020.

[32] B. Deepika, “Early prediction of chronic kidney disease by using machine learning techniques,” *Amer. J. Comput. Sci. Eng. Survey*, vol. 8, no. 2, p. 7, 2020.

[33] F. Ma, T. Sun, L. Liu, and H. Jing, “Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network,” *Future Gener. Comput. Syst.*, vol. 111, pp. 17–26, Oct. 2020.

[34] A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Aliss, “Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data,” *Sensors*, vol. 20, no. 9, p. 2649, May 2020.

[35] U. H. Amin, J. Li, Z. Ali, M. H. Memon, M. Abbas, and S. Nazir, “Recognition of the Parkinson’s disease using a hybrid feature selection approach,” *J. Intell. Fuzzy Syst.*, vol. 39, no. 1, pp. 1–21, Jul. 2020