An Optimized K-Nearest Neighbor Cross Validation Model for Enhanced Prediction of Coronary Heart Disease

1st Nicholas Mutua, 2nd Wilson Cheruyoit, 3rd Solomon Mwajele

[1]First Author, [2]Corresponding Author, [3]Corresponding Author
[1] School of Computing and Information Technology,
[1]Taita Taveta University, Mombasa, Kenya

Abstract - Coronary heart disease is a major public health issue that affects millions of people worldwide. Despite advances within computational medical treatment and prevention, identifying individuals who are at high risk for developing coronary heart disease remains a challenge due to a low accuracy rate of prediction. Machine learning algorithms, such as k-nearest neighbor, have shown promise in predicting the risk of developing coronary heart disease based on risk factors such as age, sex, smoking status, blood pressure, and cholesterol levels. However, the accuracy of these algorithms can be improved by optimizing the machine learning model classifiers using cross-validation techniques and feature selection methods to solve greater dependence on choosing the initial focal point and optimizing local minimum training. The main objective of this study was to develop an enhanced k-nearest neighbor cross validation model for enhanced prediction of heart disease. The k-Nearest Neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Index Terms – Coronary Heart Disease (CHR), k-Nearest neighbor, (KNN), Machine learning (ML), Model Classifiers (MC)

I. INTRODUCTION (HEADING 1)

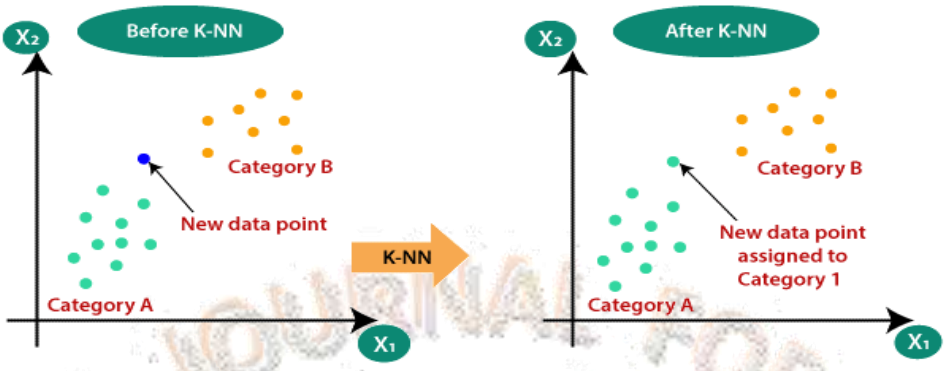
Human body is made up of various organs, all of which have their own functions. Heart is one such organ which pumps blood throughout the body and if it does not do so, the human body can have fatal circumstances [1]. One of the main reasons of mortality today is having a heart disease. The World Health Organization recommends it necessary to make sure that the cardiovascular system or any other system in the human body for that matter must remain healthy [1] . Data mining techniques can be useful in predicting heart diseases. Predictive models can be made by finding previously unknown patterns and trends in databases and using the obtained information. Data mining means to extract knowledge from large amounts of data. Machine learning is a technology which can help to achieve diagnosis of heart diseases before much damage happens to a person. As an emerging field in science and technology, machine learning can classify whether a person might be suffering from a heart disease or not [2]. Coronary heart disease is a major public health issue that affects millions of people worldwide. (Anggoro & Novitaningrum, 2021). Despite advances within computational medical treatment and prevention, identifying individuals who are at high risk for developing coronary heart disease remains a challenge due to a low accuracy rate of prediction [3].

Machine learning algorithms, such as k-nearest neighbor, have shown promise in predicting the risk of developing coronary heart disease based on risk factors such as age, sex, smoking status, blood pressure, and cholesterol levels. However, the accuracy of these algorithms can be improved by optimizing the machine learning model classifiers using cross-validation techniques and feature selection methods [4] to solve greater dependence on choosing the initial focal point and optimizing local minimum training. The research implemented a k-nearest neighbor cross-validation model that can accurately predict the risk of developing coronary hear t disease based on a set of relevant risk factors with feature selection. The model was evaluated using performance evaluation metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. The results of this study provide valuable insights into the use of machine learning algorithms for predicting the risk of coronary heart disease and inform the development of effective prevention and treatment strategies. Dealing with enormous volumes of data has become unavoidable as data quantity and rate have increased.

One definition of Big Data is the amount of information that just exceeds the capacity of technologies to store, manage, and program. Big data is assumed to be enormous, complex, and expanding from separate or unconnected sources. Big data has quickly emerged in many sectors and disciplines because of significant advancements in communication and data storage technologies, including scientific, architectural, physiological, pharmacological, and biomedical science [5].

The k-Nearest Neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point [6]. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another-NN algorithm stores all the available data and classifies a new data point based on the similarity [4]. For classification problems, a class label is assigned on the basis of a majority vote - i.e., the label that is most frequently represented around a given data point is used. While this is technically considered “plurality voting”, the term, “majority vote” is more commonly used in literature. The distinction between these terminologies is that “majority voting” technically requires a majority of greater than 50%, which primarily works when there are only two categories. It’s also worth noting that the KNN algorithm is also part of a family of “lazy learning” models, meaning that it only stores a training dataset versus undergoing a training stage [7]. This also means that all the computation occurs when a classification or prediction is being made. Since it heavily relies on memory to store all its training data, it is also referred to as an instance-based or memory-based learning method as shown in figure 1 below:

Fig.1 K-Nearest Neighbors



II. LITERATURE SURVEY

Coronary heart disease (CHD), also called coronary artery disease (CAD) and atherosclerotic heart disease, is the end result of the accumulation of atheromatous plaques5 within the walls of the arteries that supply the myocardium (the muscle of the heart) [8]. While the symptoms and signs of coronary heart disease are noted in the advanced state of disease, most individuals with coronary heart disease show no evidence of disease for decades as the disease progresses before the first onset of symptoms, often a "sudden" heart attack, finally arise [9]. [10] proposed a k-nearest neighbor (KNN) based heart disease prediction model. The author conducted an experiment to evaluate the performance of the proposed model. Moreover, the result of the experimental evaluation of the predictive performance of the proposed model is analyzed. To conduct the research, the author obtained heart disease data from Kaggle machine learning data repository.

The dataset consists of 1025 observations of which 499 or 48.68% is heart disease negative and 526 or 51.32% is heart disease positive. Finally, the performance of KNN algorithm is analyzed on the test set. The result of performance analysis on the experimental results on the Kaggle heart disease data repository shows that the accuracy of the KNN is 91.99%.
Kaggle breast cancer data repository used in this study consists of 1025 observations and 13 features. Among the 1025 observations, 499 or 48.68% are heart disease negative and 526 or 51.32% were heart disease positive.

Data analysis and prediction is required to optimize the need of necessary things. If it is required to analyzed and predict something, machine learning algorithm can be one of the best solution to deal with this type of problems [11] focused on prediction analysis using K-Nearest Neighbors (KNN) Machine Learning algorithm. Data in the dataset are processed, analyzed and predicated using the specified algorithm. Introduction of various Machine Learning algorithms, its pros and cons were discussed. The KNN algorithm with detail study was given and implemented on the specified data with certain parameters. The research work elucidates prediction analysis and explicates the prediction of quality of restaurants. achieved 74.15% accuracy. Future work can be improved by comparison with other city to get better accuracy percentage. Others prediction models can also be used to get comparative study [12].

[13] applies K-Nearest Neighbor Algorithm Analysis Application to determine the prediction of the number of Web- Based production to make it easy to predict the number of tofu production. The system functional test results show that all features in the application are able to run properly and functionally. Testing the accuracy of the prediction system K- Nearest Neighbor algorithm to determine the prediction of the number of web- based tofu production that can produces a MAPE of 0.68%
Based on implementation and testing of the prediction system for the amount of tofu production in Kedungpring District, it can be concluded as follows: Web-Based Tofu Production Prediction Application (Case Study: Kedungpring District) with the K-Nearest Neighbor method was successfully made and can run well. Web-based Tofu Production Prediction Application (Case Study: Kedungpring District) using the K-Nearest Neighbor method produces a MAPE of 0.68% and an accuracy rate of 99.32%.

Protein-protein interactions (PPIs) are an important part of many life processes in organisms. Almost all life processes are related to protein-protein interactions, and the study of protein interactions plays an important role in revealing the mysteries of life activities. In order to improve the prediction performance of protein-protein interaction, [14] based on K-Nearest Neighbor (KNN), combined with protein sequence coding methods such as Conjoint Triad (CT), Auto Covariance (AC) and Local Descriptor (LD) to construct KNN-CT, KNN-AC and KNN-LD three prediction models of PPIs.

The results show that the prediction models KNN-CT and KNN-AC have obtained accuracy rates of 94.29% and 94.69%, respectively, which are better than existing methods. The results showed that K-nearest neighbors can be a useful complement to protein-protein interactions. This experiment mainly used human protein data, which came from http://www.csbio.sjtu.edu.cn/bioinf/LR_PPI/Data.html. Among them, the positive samples were taken from the Human Protein Reference Database (HPRD, 2007 edition), and the negative samples were constructed with subcellular location information.

Most protein sequences range from 100 to 1000 in length, and the research removed protein pairs containing less than 50 residues and unusual amino acid sequences, such as B, J, O, U, X, and Z. The resulting data set contains 36,591 pairs of positive samples and 36,324 pairs of negative samples, of which 30,000 positive samples and 30,000 negative samples were randomly selected each time to form a training data set, and the rest is used as a test set to validate the model.

One of the heart attack indices that attracts many investors is the heart attack index on. One of the algorithms that can be used to predict is the k-Nearest Neighbors (kNN) algorithm. [15] used kNN regression method because it predicts numerical data. The results of the research in making the heart attack index prediction application was successfully built. The highest accuracy achieved reaches 91.81% by WSKT share.This research was conducted using Python programming language and the Flask framework. The stock data used is stock index of LQ45 property sector. They are ADHI, BKSL, BSDE, LPKR, PTPP, WIKA, and WSKT. For more information, one can access from GitHub with this link https://github.com/juliusHin/KNR_Stock_Prediction.This research using some variant values of k and used Root Mean Square Error (RMSE) and R2 (R Squared Error) to measure the accuracy level.

However, obtaining tremendous labeled data from experiments is a challenge for humans. Big data analysis has proposed some solutions to deal with this challenge. Big data technology has developed very fast and has been applied in many areas. In the bioinformatics area, big data analysis solves a large number of problems, particularly in the area of active learning. Active learning is a method of building more predictive models with less labeled data.

Active learning establishes models with less data by asking the oracle (human) for the most valuable samples to train models. Hence, active learning's application in making vaccines is meaningful that the scientists do not need to do tremendous experiments. [16] proposed a more robust active learning method based on uncertainty sampling and K-nearest density and applies it to the vaccine manufacture.

This research evaluated the new algorithm with accuracy and robustness. In order to evaluate the robustness of active learners, a new robustness index is designed in this research. And this research compared the new algorithm with a pool-based active learning algorithm, density-weighted active learning algorithm, and traditional machine learning algorithm. Finally, the new algorithm is applied to epitope prediction of B-cell data, which is significant to making vaccines.

The prediction of heart attack has become an increasingly popular research hotspot in the field of medical engineering, which benefits maritime safety supervision and security. Existing methods of heart attack prediction based on motion characteristics have a large uncertainty and cannot guarantee trajectory prediction accuracy of the heart attack. An improved method of location prediction using k-nearest neighbor (KNN) was proposed by [17].

An expanded circle area of the latest point of the target heart attack is first generated to find the reference points with similar movement characteristics in the constraints of distance and time intervals. Then, the top k-nearest neighbors are determined based on the degree of similarity. Relationships between the reference point of each neighbor and the latest points of the target ship are calculated.
The predicted location of the target ship was then be determined by a weighted calculation of the locations of all neighbors at the predicted time and their relationships with the target ship. Experiments of ship location prediction in 10 min, 20 min, and 30 min were conducted. The correlation coefficient of heart attack prediction error for the three experiments was 0.992, 0.99, and 0.9875, respectively. The results showed that ship location prediction with reference to multiple nearest neighbors with similar movements can provide better accuracy.

Although heart location recognition in a short time works with the model, it was assumed that the factors of the target ship and the similar ships retrieved in the KNN method were simple, so it is not applicable to long-term prediction. Moreover, the weight of the parameters is not dynamic in the similarity model. In future studies, the research suggests executing measures based on trajectory classification with long-distance and short-distance vessels to predict within the defined range and evaluating the prediction errors with MAE and RMSE. The contributing factors of the environment of the heart attack combined with the AIS data should also be taken into account in the prediction of heart attack.

The research by [18] predicted the future concentration values of PM10 and SO2, which are important pollutants, by using available daily records. The predictive model was implemented for Erzincan city by using a total 651 data points observed for period from 2016 through 2018. In the modelling process, data was divided into two groups; 400 the data points are utilized for training and the remaining 251 data points are used for verification.

The wavelet transform technique was combined with the K-Nearest Neighbor (KNN) method to develop a predictive model called as Wavelet- KNN approach for increasing the modeling success. In the present study, the wavelet-KNN approach is provided better prediction results compared to stand-alone KNN method. It was noted that the combination of wavelet with KNN tool was enhanced the prediction performance of model. This study showed that the KNN method is one of the simplest machine learning methods and can be used for prediction of air pollution model

METHODOLOGY

The datasets and total population used was obtained from an online freemium repository https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset. This data set dates from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 heart disease attributes.

The research was based on purposive sampling technique, a form of non-probability sampling in which decisions concerning the individuals to be included in the sample were based upon a variety of criteria which included specialist knowledge on the survey research.

Purposeful sampling is widely used in qualitative research for the identification and selection of information rich cases related to the phenomenon of interest. Although there are several different purposeful sampling strategies, criterion sampling appears to be used most commonly in implementation research [19]. This involves identifying and selecting individuals or groups of individuals that are especially knowledgeable about or experienced with a phenomenon of interest.

Despite its wide use, there are numerous challenges in identifying and applying the appropriate purposeful sampling strategy in any study. For instance, the range of variation in a sample from which purposive sample is to be taken is often not really known at the outset of a study. To set as the goal the sampling of information-rich informants that cover the range of variation assumes one knows that range of variation [19].

The final sample size from the population of the dataset is 68 kilo megabytes in size with the following final 12 attribute information from total population of 76 attributes:

i. age
ii. sex
iii. chest pain type (4 values)
iv. resting blood pressure
v. serum cholesterol in mg/dl
vi. fasting blood sugar > 120 mg/dl
vii. resting electrocardiographic results (values 0,1,2)
viii. maximum heart rate achieved
ix. exercise induced angina
x. old peak = ST depression induced by exercise relative to rest
xi. the slope of the peak exercise ST segment
xii. number of major vessels (0-3) colored by fluoroscopy

Figure 2 below show the implemented K-Nearest Neighbor Cross validation Model for Enhanced Prediction of Heart disease

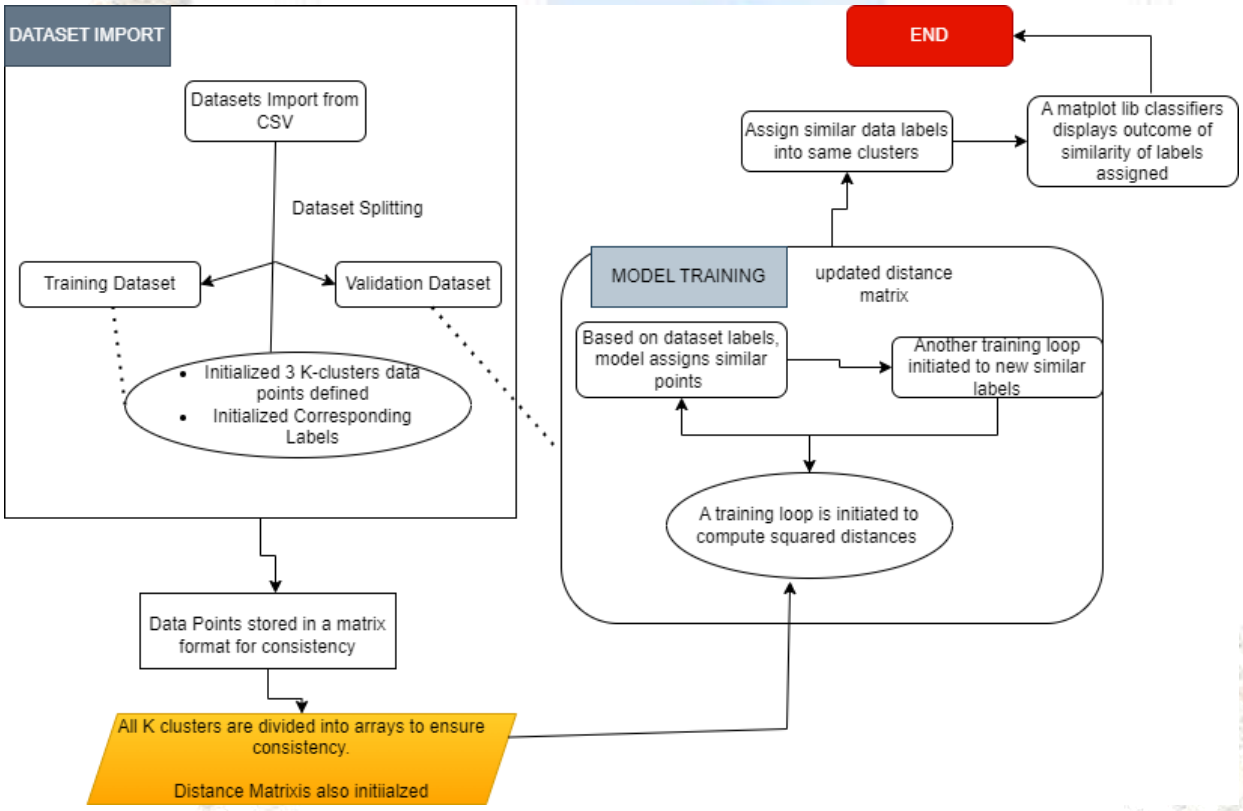


Figure 2: The implemented K-Nearest Neighbor Cross validation Model for Enhanced Prediction of Heart disease

III. DISCUSSIONS, RESULTS AND ANALYSIS

First datasets are imported into the model and divided into 80% training and 20% validation datasets. Then there is an initialization to divide datasets into cluster points with initialized labels. Then data is stored into a matrix $\tau = \{(xi, yi)\}$ n vector. The definition of matrix vectors x(1), x(2), . . . , x(n) as the featured are ordered by closeness to x in some dis- i=1 be the training set, with yi $\in \{0, \ldots, c-1\}$, and x a new feature distance dist(x, xi), the Euclidean distance $kx-x'k$. Let $\tau(x) := \{(x(1), y(1)) \ldots, (x(K), y(K))\}$ the subset of τ that contains K feature vectors xi that are closest to x.

Then the K-nearest neighbors classification rule classifies x according to the most frequently occurring class labels in τ(x). If two or more labels receive the same number of votes, the feature vector is classified by selecting one of these labels randomly with equal probability. For the case K = 1 the set τ(x) contains only one element, say (x′, y′), and x is classified as y predictions.

(1) Experiment Set-Up

The experiments were run on Google Colab due to its faster GPU processing speed with its associated libraries enabled model simulation. Python3 is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python3 build the entire structure code developed. Python3 made it easy to achieve high-level interaction nature of scientific libraries for a good platform of simulation development. This prevents the sensor nodes from discovering legitimate paths that are more than two hops away.

(2) Model Training

The heart attack.csv dataset used the features shown in the following table 4.1, with the main objective of the enhanced k-nearest model to predict whether someone will have coronary heart disease or not.

Table 4.1: Data Dictionary for the features

Feature Name	Description
Sex	Gender of a Person
CP	Chest Pain type
TRTBPS	Resting Blood Pressure (in mm Hg)
CHOL	Cholesterol in mg/dl fetched via BMI Sensor
FBS	(Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
Restecg	resting electrocardiographic results
Thalachh	maximum heart rate achieved
exng	exercise induced angina (1 = yes; 0 = no)
Age	Age of the Subject
Outcome	No Coronary Heart Attack,1-present Coronary Attack

The total data was split into train and test objects, x_train and y_train, to develop a trained model object. The k-nearest model vector classifier rule classifies x according to the most frequently occurring class labels in τ(x). If two or more labels receive the same number of votes, the feature vector is classified by selecting one of these labels randomly with equal probability parameters accounted for fine-tuning a model to get the best version of the model.

The trained model objects are then stored as kst. Using the same object, accuracy for the model was generated. The base accuracy was 82.9%. The best model accuracy being 82.9%, is saved as best for coronary heart attack prediction which is uploaded to the validation datasets to generate predictions and use it for populating final results. The model was then deployed.
For the person at record no. 8.2, the model predicted coronary heart attack and the actual is also coronary heart attack. The cholesterol level was considered to be the most important feature, followed by Resting Blood Pressure, age, and Chest pain Type as shown in the following Table 4.2.

Table 4.2. Changes in Feature Values Impacts Prediction

Feature Name	Old_value	New_value
CP (Chest **Pain Rate)**	4	7
TRTBPS (Blood Pressure)	45	38
CHOL (cholesterol)	127	95
FBS (Blood Sugar)	0.55	0.67
Thalachh (Heart rate)	24/3 Per Min	27/5 Per Min
Exng (Exercise)	0.56	0.44
Age	28	55

(3) Data Table Results

The confusion Matrix after a series of deployments of simulations was $\begin{bmatrix} 98 & 5 \\ 7 & 89 \end{bmatrix}$. The following table 4.1 shows the detailed confusion matrix results for the enhanced KNN model for predicting heart disease.

Table 4.1: Confusion Matrix Results

	Positive	Negative	
Positive	98	5	**Sensitivity = 93.3%** $\frac{True\ Positive}{True\ Positive + False\ Negative}$
Negative	7	89	**Specificity = 94.68%** $\frac{True\ Negative}{True\ Negative + False\ Positive}$
	Precision = 95.15% $\frac{True\ Positive}{True\ Positive + True\ Negative}$	**Negative Predictive Value = 7.25%** $\frac{True\ Negative}{True\ Negative + False\ Negative}$	**Accuracy = 93.97%** $\frac{True\ Positive + True\ Negative}{TP + TN + FP + FN}$

Accuracy was the ratio of correct predictions to the total number of predictions. Precision was the ratio of true positives to the total number of positive predictions. Sensitivity (Recall or True Positive Rate) was the ratio of true positives to the total number of actual positives. Specificity (True Negative Rate) was the ratio of true negatives to the total number of actual negatives. Negative Error Rate (False Negative Rate) was the ratio of false negatives to the total number of actual negatives. Our enhanced KNN showed a good performance through the simulations from all the evaluation metrics.

IV. CONCLUSIONS

The k-nearest model vector classifier rule classifies x according to the most frequently occurring class labels in τ(x). If two or more labels receive the same number of votes, the feature vector is classified by selecting one of these labels randomly with equal probability parameters accounted for fine-tuning a model to get the best version of the model. The simulation results were presented in a confusion matrix with an accuracy of 93.7% which was the ratio of correct predictions to the total number of predictions and a precision of 94.5% which was the ratio of true positives to the total number of positive predictions. Future research work can enhance the algorithm by using dynamic dataset to quantify the accuracy and improve the classifiers.

V. REFERENCES

[1] M.-H. Biglu, M. Ghavami, and S. Biglu, "Cardiovascular diseases in the mirror of science," *J. Cardiovasc. Thorac. Res.*, vol. 8, no. 4, pp. 158–163, 2016, doi: 10.15171/jcvtr.2016.32.

[2] N. Sharma, A. Raj, V. Kesireddy, and P. Akunuri, "Machine Learning Implementation in Electronic Commerce for Churn Prediction of End User," no. 5, pp. 20–25, 2021, doi: 10.35940/ijsce.F3502.0510521.

[3] K. Lalitha and S. Murugavalli, "A survey on image retrieval techniques," *Adv. Parallel Comput.*, vol. 37, no. 1, pp. 396–400, 2020, doi: 10.3233/APC200174.

[4] M. Suyal and P. Goyal, "A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning," *Int. J. Eng. Trends Technol.*, vol. 70, no. 7, pp. 43–48, 2022, doi: 10.14445/22315381/IJETT-V70I7P205.

[5] S. S. Mesakar and M. S. Chaudhari, "A Review of Clustering Algorithms 1 1,2," vol. 8491, no. October, pp. 4–6, 2013, doi: 10.5281/zenodo.7243829.

[6] T. Oladipupo, "Machine Learning Overview," *New Adv. Mach. Learn.*, no. February 2010, pp. 8–18, 2010, doi: 10.5772/9374.

[7] C. S. Ingulkar and A. N. Gaikwad, "Hand Data Glove: A wearable real time device for human computer Interaction," *Int. J. Sci. Eng.*, vol. 1, no. 2, pp. 99–104, 2013.

[8] O. Stephen, "The Study of the Application of Data Encryption Techniques in Cloud Storage to Ensure Stored Data Integrity and Availability," *Int. J. Sci. Res. Publ.*, vol. 4, no. 10, pp. 1–7, 2014, [Online]. Available: www.ijsrp.or.

[9] A. Garg, B. Sharma, and R. Khan, "Heart disease prediction using machine learning techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, 2021, doi: 10.1088/1757-899X/1022/1/012046.

[10] T. A. Assegie, "Heart disease prediction model with k-nearest neighbor algorithm," *Int. J. Informatics Commun. Technol.*, vol. 10, no. 3, p. 225, 2021, doi: 10.11591/ijict.v10i3.pp225-230.

[11] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System model for prediction analytics using /(-nearest neighbors algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019, doi: 10.1166/jctn.2019.8536.

[12] B. Tekaya, S. El Feki, T. Tekaya, and H. Masri, "Recent applications of big data in finance," *ACM Int. Conf. Proceeding Ser.*, no. December 2020, 2020, doi: 10.1145/3423603.3424056.

[13] M. Munif, M. Mustain, and K. Yahya, "Analysis of the K-Nearest Neighbor Algorithm to Determine the Prediction of Tofu Production," *Appl. Technol. Comput. Sci. J.*, vol. 5, no. 1, pp. 57–64, 2022, doi: 10.33086/atcsj.v5i1.3677.

[14] Y. Gui and X. Wang, "Application of K-nearest neighbors in protein-protein interaction prediction," *Highlights Sci. Eng. Technol.*, vol. 2, pp. 125–131, 2022, doi: 10.54097/hset.v2i.564.
[15] J. Tanuwijaya and S. Hansun, "LQ45 stock index prediction using k-nearest neighbors regression," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 2388–2391, 2019, doi: 10.35940/ijrte.C4663.098319.
[16] T. Lu, "K-Nearest Robust Active Learning on Big Data and Application in Epitope Prediction," *Wirel. Commun. Mob. Comput.*, vol. 2021, no. November, 2021, doi: 10.1155/2021/8752022.
[17] M. Zhang, L. Huang, Y. Wen, J. Zhang, Y. Huang, and M. Zhu, "Short-Term Trajectory Prediction of Maritime Vessel Using k-Nearest Neighbor Points," *J. Mar. Sci. Eng.*, vol. 10, no. 12, 2022, doi: 10.3390/jmse10121939.
[18] A. ALTUNKAYNAK, E. E. BAŞAKIN, and E. KARTAL, "Dalgacik K-En Yakin KomşulukYöntemiİle HKirliliğiTahmini," *Uludağ Univ. J. Fac. Eng.*, no. December, pp. 1547–1556, 2020, doi: 10.17482/uumfd.809938.
[19] L. A. Palinkas, S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood, "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research," *Adm. Policy Ment. Heal. Ment. Heal. Serv. Res.*, vol. 42, no. 5, pp. 533–544, 2015, doi: 10.1007/s10488-013-0528-y.