© February 2022| IJIRT | Volume 8 Issue 9 | ISSN: 2349-6002

Prediction of Diabetes Disease Using Classification Techniques

Peddineni Kalpana[1], Dr. B V V Siva Prasad[2]
[1]*Research Scholar, Career Point University, Kota, Rajasthan*
[2]*Research Supervisor, Career Point University, Kota, Rajasthan*

***Abstract* - Data mining is computational procedure of observing models in generous information sets counting schedules at connection reason behind man-made intellectual ability, AI, experiences, and information base systems. Various information mining procedures have been utilized by experts in finding of Diabetes illness. However applying information mining is valuable to medical services, sickness determination, and therapy, few explores have examined creating treatment plans for patients. The fundamental issue in the diabetes information grouping is just due to in adequate assets and information legitimate mining has not been finished. To eliminate the issue of the information mining in medical care appropriate information peculiarities must be pre-handled and excess should be eliminated from the dataset. This paper presents various classification techniques for the prediction of diabetes in the filed of data mining and also it gives predictive analysis with various data mining techniques.**

***Index Terms* - diabetes disease, classification techniques, data mining.**

I.INTRODUCTION

Diabetes is a disease that occurs when the body's insulin production is impaired or when the body's insulin is not used properly, resulting in high blood glucose levels. Food is broken down into glucose by body cells, and this glucose must be transported to all of the body's cells. Insulin is a synthetic that arranges the glucose that is transported into the body cells by separating the meal. Any alteration in insulin production causes an [1] increase in glucose levels, which can cause damage to the tissues and malfunction in the organs. When glucose levels are higher than usual (4.4 to 6.1 mmol/L), an individual is considered to have diabetes. Type 1, Type 2, and Gestational diabetes are the three main types of diabetes.

The three types of diabetes are depicted as follows:

1. Type 1 diabetes - Despite the fact that only around 10% of diabetes patients have [2-3] this type of diabetes, the number of cases has recently increased in the United States. The condition manifests as a resistant framework disorder that begins while a person is under the age of 20 years old, also known as young adult diabetes. The pancreatic cells that create insulin have been crushed by the body's security game plan in this type of diabetes. Patients with Type 1 diabetes should take insulin injections along with regular blood tests and dietary restrictions.
2. Type 2 — This type accounts for over 90% of all diabetes occurrences and is also known as adult-onset diabetes or non-insulin-dependent diabetes. For the time being, [4-6] numerous parts of the body have become insulin-safe, which has increased interest in insulin. The pancreas no longer produces a significant amount of insulin. To keep track of this type of diabetes, patients must adhere to a strict dietary regimen, exercise regularly, and check their blood glucose levels. Bulkiness, being overweight, and being physically sedentary are all factors that might lead to diabetes. Furthermore, as one grows older, the danger of having diabetes is considered as an added concern. Peripheral diabetes, also known as Pre-Diabetes, affects a larger percentage of Type 2 diabetes patients. It is a disease in which blood glucose levels are higher than normal but not as high as in diabetic individuals.
3. Gestational diabetes - is a kind of diabetes that develops in pregnant women as a result of high blood sugar levels caused by the pancreas' inability to deliver an adequate [7-9] amount of insulin. Choosing not to receive therapy might lead to complications at work. This kind of diabetes may be managed by following a strict diet and using insulin. This vast array of diabetes types is extremely dangerous and necessitates treatment; yet, if detected early enough, the dangers associated with them can be avoided.

II. DATA MINING

Data mining is the extraction of nontrivial and important data from huge data vaults to track down unpretentious patterns and connections in the information concealed to the unaided eye. To track down these patterns also connections, the information mining process utilizes AI calculations and measurements to process the information. Information mining is a critical piece of a more extensive interaction known as Knowledge Discovery (KD). KD is formed by a few stages that include: (I) information preprocessing, that cleans, chooses and changes the dataset, (ii) information mining strategies to separate information designs, (iii) design assessment where the examples found before are considered lastly (iv) information show, where the last ends and designs are shown. One of the principle objectives of information mining processes is to track down information relations in a little subset of information and transform it into a common principle to be applied in concealed datasets. Information mining depends on some general ideas. An information object is a substance that might be characterized as an occasion. An case is ordinarily portrayed by characteristics (or highlights) and an information tuple is an information object distinguished by a key (class) and portrayed by a bunch of quality qualities. A total arrangement of examples is a dataset. A dataset can be envisioned as a table where lines compare to cases and the segments to the properties. A property is an information field, addressing an attribute of a case. Ascribes can be ostensible, while including downright qualities; parallel, which is an ostensible characteristic with just two states: 0 or 1 (bogus or valid); or numeric, having quantifiable amounts addressed in whole number or genuine values. The order task present in each datum mining application comprises in separating models to portray information classes. The arrangement has two stages: the learning step, order model is built, grouping step, model is applied into another dataset to anticipate class marks. At the point when the class mark is referred to, the learning step is characterized as administered getting the hang of, appearing differently in relation to solo learning (or grouping), when nor the class name or the quantity of classes are obscure. Datasets and occasions might have a few issues that must be tended to before the order task. Occasions may astoundingly go astray from the standard qualities and conduct, including exceptions and are normally portrayed as commotion. Class unevenness if not tended to accurately may lead to overfitting. Overfitting happens when the learning model contains decides that are simply explicit to that dataset and not appropriate to the overall model, in this manner compromising the speculation used to construct the model.

Information mining is the computational technique of tracking down models in generous datasets including schedules. Adjacent to the rough examination step, it incorporates data set, data organization points of view, pre-handling, acceptance thoughts, nature examinations, post-planning of tracked down constructions, representation and web updating. It is like manner is a famous articulation and is routinely associated with any kind of broad scale data or information taking care of and furthermore any utilization of PC decision genuinely steady organization, including automated thinking, AI, and business information. The genuine task is the modified considerable measures of information to focus as of now dark entrancing models, social events, phenomenal data (peculiarity acknowledgment).

III. MACHINE LEARNING CLASSIFICATION TECHNIQUES

Gathering is one kind of insightful illustrating. Even more expressly, gathering is course of delegating things to predefined arrangements: Given a lot of named records, collect a model for instance, a decision tree, and expect marks for future unlabeled records Model construction in the gathering system is a controlled learning issue. Getting ready models are portrayed the extent that (1) attributes, which can be unmitigated—i.e., unordered meaningful characteristics—or numeric; and (2) class name, which is moreover called the expected or result property. If the last choice is unmitigated, then, we have a request issue. Expecting the last choice is numeric, then, we have a backslide issue. The arrangement models are taken care of using some AI estimation to manufacture a decision limit, for instance, a decision tree to predict signs of new data. Data Mining methods, for instance, portrayal, association and packing are generally used to eliminate the concealed, in advance hid data from voluminous of informational indexes. Of the diverse data examination technique, course of action is a controlled AI strategy which makes gauges about the future class models by arranging instances of testing data to the predefined class marks which is gained from the furnished instances of classes with class names. There are a couple of models in groupings, for instance, probabilistic model, groundbreaking algorithmic model, etc Grouping involves expecting a particular outcome reliant upon a given information. To expect the outcome, the estimation cooperation a planning set, containing a lot of attributes and the different outcome, ordinarily called objective or assumption quality. The estimation endeavors to track down associations between the credits, which would expect the outcome. In estimation, a given enlightening assortment which isn't seen before is called conjecture set, which contains comparable game plan of attributes, except for the assumption characteristic, which are not yet known. Computation examination the data, and produces an estimate. The accuracy of figure is described by the tolerability of the estimation used. Course of action methodology is prepared for taking care of a more broad collection of data, than backslide and it is fostering its fame. The goal of a classifier isn't to research the data to track down charming sections yet rather to finish up how new records should be gathered. Gathering plans for data mining furthermore use a grouping of estimations.

- SVM

There is a ton of authentic procedures, which target handling equal request endeavors, for instance, the assessment of the credit surviving from adventures. The most famous strategies unite standard quantifiable frameworks like direct Discriminant Analysis and non-parametric genuine models. SVMs are another promising non-quick, course of action method, which as of late showed unbelievable outcomes in the clinical diagnostics, optical individual assertion, electric weight choosing and different fields. Applied to dissolvability examination, the ordinary objective of this enormous number of portrayal is to encourage limit, which can definitively disconnect dissolvable, bankrupt associations. The score reduces the information included in an association's accounting report to a one-layered abstract pointer, which is a portion of certain markers, primarily monetary in nature. SVMs, which are connected to and include components of non-parametric applied estimates, neural connections, and AI, are another technique useful for equal portrayal tasks. SVMs, like traditional methods, classify an organization as dissolvable or destroyed based on its score respect, which is a set of money-related limits. This limit, however, is neither linear nor parametric. The fundamentals of SVMs will be swiftly introduced in this manner. To illustrate edge support in a dealt with context, an example of a direct SVM will be presented first, where the score work is still quick and parametric. By injecting a bit, the SVM will be created non-straight and non-parametric after a short period of time. As previously stated, it is this brand identity that distinguishes SVMs as a useful tool for credit scoring in situations when open data cannot be obtained or its relationship with the PD is non-rambling. Another feature of dissolvability testing is the ability to compare various scores with the related probability of default (PD) over a certain time period. This perspective is especially important in the Euro structure, where recognize scoring is used to orchestrate the capacity of association recognize liabilities as a safeguard for public bank revaluation activities, because capability is linked to a benchmark regard comparable to the annual PD. The assurance of a credit scoring game plan method is a challenging issue, because a valid option based on open data might aid chip away at the precision in credit scoring practise. Of course, this selection should not be viewed as a "either/or" choice, since multiple depiction strategies may be combined, resulting in the development of a comprehensive credit rating system. SVMs are provided as a feasible gathering approach for credit rating in the beginning of the study. Vapnik was the first to propose the Support Vector Machine (SVM), which has sparked significant interest in the AI research community. A few recent studies have shown that SVM (support vector machines) are capable of delivering higher execution in terms of depiction correctness than other data collecting estimates. Sims have been utilised in a variety of real-world situations, including text requests, interpreted digit affirmation, tone affirmation, image collection and article area, smaller than expected group quality verbalization data evaluation, and data representation. Sims has been found to perform consistently better than other coordinated learning processes. In any event, the performance of SVM is highly dependent on how the cost and piece restrictions are chosen for specific datasets. As a result, determining the best limit setting usually necessitates extensive cross endorsement by the consumer. For the most part, this message is interpreted as model assurance. One reasonable objection to model assurance is that the link is rather dull. We looked at a number of routes to see if there were any limitations to the use of the SVM estimation that may affect the findings. These constraints include the number of planning models, the standard deviation of the Gaussian piece, relative loads associated with slack components to handle non-uniform data transit, and the number of spot works. For request and backslide, VMs are a collection of linked directed learning algorithms. They have a site with a collection of straight representation that has been summarized. SVM has the unique virtue of limiting specific depiction mix-up while still assisting the numerical advantage. As a result, SVM is also known as Maximum Margin Classifiers. SVM is based on the Structural Risk Minimization principle (SRM). A maximum disconnecting hyperplane is produced by mapping the input vector to a higher layered space using SVM. On either side of the hyperplane, two comparable hyperplanes are used to diversify the data. The segregating hyperplane is a hyperplane that aids in the separation of two comparable hyperplanes. The notion is that the bigger the edge or distance between these similar hyperplanes, the better the classifier's hypothesis will be screwed up. (x1,y1), (x2,y2), (x3,y3), (x4,y4).........., (xn, yn) are the main components of the building. Where yn=1/ - 1 denotes the class in which that particular point xn belongs. n stands for the number of tests. Each x n is a real p-layered vector. To prepare for variables (credits) with more contrast, scaling is required. Through the dividing (or secluding) hyperplane, which accepts w. x + b = o, we may observe this Training data. Where w is a p-layered Vector and b is a scalar. The vector w coordinates inverse toward the separating hyperplane. Adding the offset limit b grants us to grow the edge. Missing b, the hyperplane is compelled to go through the start, restricting the plan. As we are charming in the best edge, we are captivated SVM and the equivalent hyperplanes. Equivalent hyperplanes can be depicted by condition

w.x + b = 1

Assuming that the planning data can be separated in a straight line, we may choose these hyperplanes such that there are no concentrations between them, and then try to increase their distance afterwards. The distance between the hyperplane is 2/w, as calculated using arithmetic. As a result, we definitely wish to limit w.

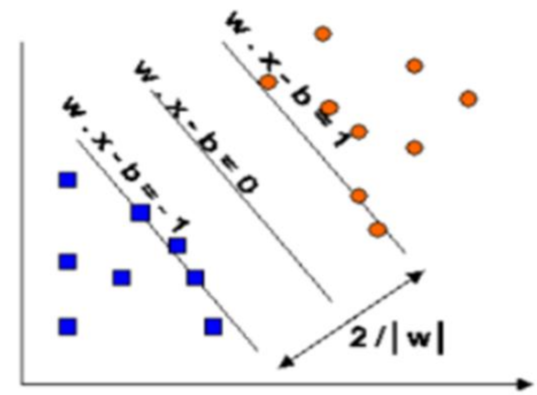

Fig 1: Margin hyperplanes for SVM

- Decision tree

Generally, a decision tree is a kind of graph that chooses a specific course of action of exercises. Particularly like a common tree, it moreover has various branches. Every decision tree appendage tends to a substitute outcome or possible reaction to an issue. Furthermore, the furthermost pieces of this tree address the fulfillment results. These trees have remarkable significance autonomous way. They give all of the potential outcomes at whatever point to envision all of the conditions. Decision trees in like manner enable the shot at exploring the result or consequence of any decision. A decision tree regularly addresses a flowchart structure. Here, every inside center point looks at to a test subject to a part, and each leaf center point imparts a class mark or a decision to make later the estimation of the general huge number of components. The branches, in any case, address a blend of features inciting the class names. Likewise, the gathering rules are the ways going from root to leaf. Utilizing decision trees is one of the strategies from estimations, data mining, and AI. These are a kind of managed AI. Here, the data assessment method isolates the data into various potential components concerning a specific limit. There are two essential substances in a decision tree. These components are centers and leaves. The oversaw learning strategy uses the decision tree as a farsighted model to explore a thing's insight in the branches to close the thing's true worth in the leaves. The decision center points address the splitting of data, and the leaves address the outcomes. The decision tree generally addresses human-like figuring attributes to make a canny decision. Thusly decision trees are adequately legitimate. Furthermore, there is a tree-like development behind the decision tree, so the reasoning is adequately comprehensible. To make it more sensible, coming up next are the phrasings to consider. Root Node: This is where the decision tree starts. The root center point tends to the all out dataset what parcels into no less than two commensurate sets. Leaf Node: The leaf centers address the last outcome. The computation can't separate the tree further resulting to coming to the leaf center point. Separating: This term infers apportioning the root center or the decision center point into different sub-centers according to the key conditions. Pruning: It is the most well-known method of slashing down the extra branches from the tree. It helps in wrapping up a ton prior by excepting inconsequential branches. Parent and Child center point: The parent center point is the tree's root center, while various center points are kid centers of the parent. Branch/Sub-Tree: The splitting of the crucial tree outlines new subtrees and branches. Decision trees are the non-parametric sort of coordinated learning.

- KNN algorithm

The inspiration driving the k Nearest Neighbors (kNN) computation is to use an informational collection in which the data centers are disengaged into a couple of isolated classes to anticipate the plan of another test point. Accept a bank has an informational index of people's nuances and their FICO score. These nuances would probably be the person's money related traits, for instance, the sum they get, whether or not they own on the other hand rent a house, and so forth, and would be used to work out the person's FICO evaluation. Nevertheless, the cycle for registering the financial assessment from the singular's nuances is exorbitant, so the bank should find some strategy for diminishing this cost. They comprehend that by the genuine thought of a FICO assessment, people who have equivalent financial nuances would be given relative FICO ratings. Subsequently, they should have the choice to use this current informational index to anticipate one more customer's FICO appraisal without playing out all of the calculations.

The calculation can be summed up as:

1. A positive whole number k is determined, alongside another example
2. We select the k passages in our data set which are nearest to the new example
3. We track down the most widely recognized arrangement of these passages
4. This is the order we provide for the new example

K Nearest Neighbor (KNN) is natural to comprehend and a simple to execute the calculation. Novices can dominate this calculation even in the beginning stages of their Machine Learning studies.

This KNN article is to:

- Comprehend K Nearest Neighbor (KNN) calculation portrayal and forecast.
- See how to pick K worth and distance metric.
- Required information planning techniques and Pros and cons of the KNN calculation.
- Pseudocode and Python execution.

K Nearest Neighbor estimate falls under the category of Supervised Learning and is commonly used for gathering and backslide. It's a flexible estimate method that may also be used to credit missing attributes and resample datasets. As the name implies, it considers how K Nearest Neighbors (Data characteristics) predict the class or consistent motive for the new Data point.

- Logistic Regression

Strategic relapse is a method of obtaining information that is used to make decisions on a certain course of action for a group of classes. Unlike direct backslide, which produces a stable quantity of characteristics, determined backslide alters its consequence by employing the essential sigmoid ability to return a probability respect, which may then be wished to anywhere between two discrete classes. Given data on time spent thinking about and test scores. Direct Regression and key backslide can anticipate different things: Direct Regression could help us with predicting the understudy's grade on a size of 0 - 100. Direct backslide conjectures are persevering (numbers in a span). Key Regression could help use with predicting whether the understudy passed or failed. Key backslide gauges are discrete (simply express characteristics or orders are allowed). We can similarly see probability scores stowed away the model's groupings. To design expected characteristics to probabilities, we use the sigmoid limit. The limit maps any authentic worth into another value some place in the scope of 0 and 1. In AI, we use sigmoid to design gauges to probabilities. For the accompanying, let all information vectors xi contain an extra part 1. This will work with documentation in permitting us to compose a basic spot item x*x for a direct mix of vector parts rather than the more unwieldy x* x + x0.By and large, a strategic relapse model works out the class participation likelihood for one of the two classifications in the informational index:

$$p(x) = 1/(1 + e^{-x*x})$$

The hyperplane of all focuses x fulfilling the condition x*x= 0 structures the choice limit between the two classes; these are the focuses for which p(x*x) =0.5. A strategic relapse model that incorporates just the first covariates is known as a fundamental impacts model; counting connection terms, for example, items makes the model nonlinear in the covariates, and subsequently more adaptable. Albeit higher adaptability might be attractive by and large, it conveys with it a higher danger for model overfitting ("retaining the preparation cases"), which can conceivably diminish a models exactness on already concealed cases. In prescient demonstrating, fitting the preparation cases is simply aspect of the assignment: effectively arranging new cases is the main goal. In strategic relapse, the model intricacy is as of now low, particularly when no or not many connection terms and variable changes are utilized. Overfitting is less of an issue for this situation. Performing variable determination is a method for diminishing a models intricacy and therefore decline the danger of overfitting. As referenced previously, this might cause a misfortune in the models adaptability.

- Random Forest

A controlled AI computation based on decision tree estimations is known as a Random Forest. This estimation is used in a variety of industries, such as banking and online commerce, to predict lead and outcomes. Irregular Forest is an AI system that deals with difficulties like backslides and course of action. It employs group acknowledgment, which is a method that combines many classifiers to provide answers to complicated problems. Various decision trees are used in an uncertain forest estimate. Through pushing or bootstrap conglomerating, the 'forest area' formed by the subjective boondocks evaluation is ready. Firing is a type of group meta-computation that reduces the accuracy of AI predictions. The (unpredictable forest) estimate determines the outcome based on the decision tree assumptions. It forecasts by averaging or averaging the results from various trees. The precision of the outcome is determined by the number of trees planted. The restrictions of a decision tree estimation are obliterated by a subjective woodlands. It reduces overfitting in datasets and assembles precision. It calculates estimates without requiring many game plans to be played in groups. The building squares of a subjective forest area estimation are decision trees. A decision tree is a decision-making aid that depicts a strategy in the form of a tree. A decision tree diagram will help us understand how subjective forest area computations function. Decision centres, leaf centre points, and a root centre point are the three portions of a decision tree. A decision tree estimate divides a readiness dataset into branches, which are then divided further into different branches. This process is repeated until a leaf centre point is reached. The middle point of the leaf cannot be separated any farther. The decision tree's centre points address the credits that are utilised to predict the result. The decision centre points are linked to the leaves. In a decision tree, the going with chart depicts the three types of centres.

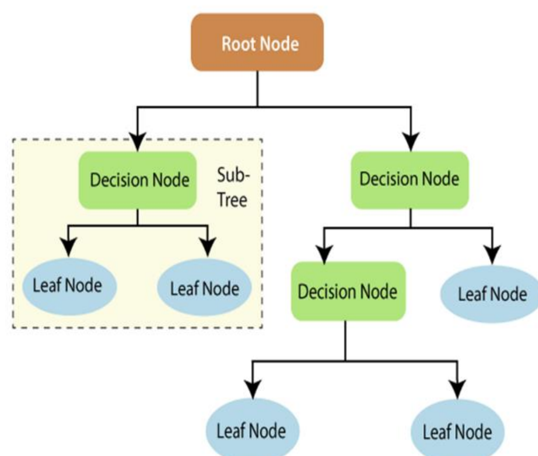

Fig 2: Flowchart of Random Forest

More information on how decision trees function may be found in information theory. The building squares of decision trees are entropy and information gain. Our comprehension of how decision trees are gathered will be harmed by a blueprint of these primary ideas. Entropy is a measure of how vulnerable something is. Given a large number of free factors, information gain is a measure of how much the actual variable's weakness is reduced. Using free factors (features) to secure knowledge about a desired variable is part of the information gain concept (class). The information gain is calculated using the entropy of the objective variable (Y) and the prohibitive entropy of Y (given X). For the time being, the unanticipated entropy is subtracted from Y's entropy. In the arrangement of decision trees, information gain is utilised. It aids in the reduction of tree weakening. A large information gain indicates that a significant amount of weakness (information entropy) has been eliminated. Partitioning branches, which is a big element of improving decision trees, requires a lot of entropy and information gain. Setting up root centres and confining centres is done discretionarily in the last choice, which is the main difference between decision tree estimate and self-assertive woodlands computation. The essential estimate is created by the discretionary woods using the stowing process. Rather than employing simply one model, firing integrates the use of several instances of data (getting ready data).

IV. CONCLUSION

Various methodologies and information mining systems utilized to figure different diabetes problems at a beginning phase are discussed in this paper. By the utilization of information mining apparatuses and cycles, diabetes is stayed away from and treatment rates are diminished. For the most part many tests are done that include grouping or order of huge scope information. Anyway numerous tests could entangle the fundamental analysis cycle and lead to trouble in receiving last products, especially in the situation where many tests are performed. The tip of this work is to dissect presentation of different arrangement methods for a bunch of enormous information.

REFERENCE

[1] D. S. Sisodia and R. Agrawal, "Data Imputation-Based Learning Models for Prediction of Diabetes," 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 966-970.

[2] R. C. Anirudha, R. Kannan and N. Patil, "Genetic algorithm based wrapper feature selection on hybrid prediction model for analysis of high dimensional data," 2014 9th International Conference on Industrial and Information Systems (ICIIS), 2014, pp. 1-6, doi: 10.1109/ICIINFS.2014.7036522.

[3] R. M. Khalil and A. Al-Jumaily, "Machine learning based prediction of depression among type 2 diabetic patients," 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 2017, pp. 1-5, doi: 10.1109/ISKE.2017.8258766.

[4] Z. T. Al-Ars and A. Mahdi Aldabbagh, "Predicting the Early Re-admission of Diabetic Patients Using Different Data Mining Techinques," 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2021, pp. 1-8, doi: 10.1109/ICECCT52121.2021.9616746.

[5] D. Jayakumar, A. Elakkiya, R. Rajmohan and M. O. Ramkumar, "Automatic Prediction and Classification of Diseases in Melons using Stacked RNN based Deep Learning Model," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262414.

[6] M. NirmalaDevi, S. A. alias Balamurugan and U. V. Swathi, "An amalgam KNN to predict diabetes mellitus," 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), 2013, pp. 691-695, doi: 10.1109/ICE-CCN.2013.6528591.

[7] U. Ojha and S. Goel, "A study on prediction of breast cancer recurrence using data mining techniques," 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017, pp. 527-530, doi: 10.1109/CONFLUENCE.2017.7943207.

[8] R. G. Franklin and B. Muthukumar, "Survey of Heart Disease Prediction and Identification using Machine Learning Approaches," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 553-557, doi: 10.1109/ICISS49785.2020.9316119.

[9] X. Wang, Y. Lu and W. -B. Chen, "Promote Retinal Lesion Detection for Diabetic Retinopathy Stage Classification," 2020 IEEE Conference on Multimedia Information

Processing and Retrieval (MIPR), 2020, pp. 31-34, doi: 10.1109/MIPR49039.2020.00014.

[10] B. V. Sumana and T. Santhanam, "Prediction of diseases by cascading clustering and classification," 2014 International Conference on Advances in Electronics Computers and Communications, 2014, pp. 1-8, doi: 10.1109/ICAECC.2014.7002426.

[11] R. C. Anirudha, R. Kannan and N. Patil, "Genetic algorithm-based wrapper feature selection on hybrid prediction model for analysis of high dimensional data," 2014 9th International Conference on Industrial and Information Systems (ICIIS), 2014, pp. 1-6, doi: 10.1109/ICIINFS.2014.7036522.

[12] D. S. Sisodia and R. Agrawal, "Data Imputation-Based Learning Models for Prediction of Diabetes," 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 966-970, doi: 10.1109/DASA51403.2020.9317070.

[13] K. Rajmohan, I. Paramasivam and S. Sathyanarayan, "Prediction and Diagnosis of Cardiovascular Disease -- A Critical Survey," 2014 World Congress on Computing and Communication Technologies, 2014, pp. 246-251, doi: 10.1109/WCCCT.2014.74.

[14] Y. K. Gupta and S. Kumari, "Performance Evaluation of Distributed Machine Learning for Cardiovascular Disease Prediction in Spark," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 1506-1512, doi: 10.1109/ICOEI51242.2021.9452955.

[15] D. Deka, J. P. Medhi and S. R. Nirmala, "Detection of macula and fovea for disease analysis in color fundus images," 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), 2015, pp. 231-236, doi: 10.1109/ReTIS.2015.7232883.

[16] A. Azari, V. P. Janeja and A. Mohseni, "Predicting Hospital Length of Stay (PHLOS): A Multi-tiered Data Mining Approach," 2012 IEEE 12th International Conference on Data Mining Workshops, 2012, pp. 17-24, doi: 10.1109/ICDMW.2012.69.