A Novel Approach for Forecasting Disease Using Machine Learning

Jeevitha D[1], Tamilakilan A[1], Mr. G. Sudhakar[2], Dr. S. Nithiyanandam[3]

[1]UG Student, Department of Computer Science Engineering, Sri Sai Ranganathan Engineering, Coimbatore, Tamilnadu, India- 641109

[2]Assistant Professor, Department of Computer Science Engineering, Sri Sai Ranganathan Engineering, Coimbatore, Tamilnadu, India- 641109

[3]Professor, Department of Mechanical Engineering, Jayshriram Engineering College Tiruppur

--***--

__Abstract__- The wide variation of PC based innovation in the medical services industry brought about the collection of electronic information. Because of the significant measures of information, clinical specialists are confronting difficulties to investigate side effects precisely and recognize infections at a beginning phase. Nonetheless, managed AI (ML) calculations have displayed critical potential in astounding standard frameworks for infection analysis and helping clinical specialists in the early recognition of high-risk disease. In this writing, the point is to perceive patterns across different kinds of regulated ML models in sickness location through the assessment of execution measurements. The most noticeably examined regulated ML calculations were Naïve Bayes (NB), Decision Trees (DT), K-Nearest Neighbor (KNN). According to discoveries, K-Nearest Neighbor (KNN) is the most sufficient at distinguishing kidney infections, Parkinson's sickness and Heart disease. At long last, Logistic Regression (LR), and Convolutional Neural Networks (CNN) anticipated in accuracy bosom disease and normal sicknesses, separately.

__Keywords__—Health Care, Supervised Machine Learning, Disease Prediction, Common Disease, Chronic Disease, Heart Disease, Breast Disease, Parkinson Disease

I. INTRODUCTION

1.1 Inspiration

The rise of Artificial Intelligence (AI) empowered electronic frameworks to see, think and work in a shrewd way like people [1]. Artificial intelligence is a multidisciplinary idea of ML, Computer Vision, Deep Learning, and Natural Language Processing [2]. ML calculations apply different improvement, factual, and probabilistic methods to gain from information that was created from previous encounters, and convey it in decision making [3]. These calculations considered to be applied in many disciplines including network interruption acknowledgment, client buy conduct identification, process fabricating streamlining, Mastercard misrepresentation location, and sickness balance. A significant number of these applications have been planned utilizing the regulated learning approach. In this methodology, datasets with realized names are initiated to forecast models to foresee unlabelled models [2], [3]. This presents the theory that clinical specialists can use managed advancing as an incredible asset to direct disease determination all the more productively [4].

Medicaid administrations and communities for Medicare announced that half of Americans had numerous chronic disease, which drove the US medical care to spend around $3.3 trillion out of 2016 that adds up to $10,348 per individual in the US [5]. Additionally, the World Health Organization and World Economic Forum revealed that India had an enormous deficiency of $236.6 billion by 2015 in light of deadly infections, brought about by hunger and bleak ways of life [6]. Such uses uncovered how inclined individuals are to a range of disease, which exhibited that it is so fundamental to identify infections ahead of schedule, to lessen the casualty of these disease therefore. Furthermore, early infection forecast can reduce the monetary tension on the economy and guarantee better upkeep on the general prosperity of the local area [5],

As per Yuan [7], ML calculations are exceptionally defenseless to mistakes due to two variables. Right off the bat, it relies upon the quality and the determination of the datasets, which is critical to accomplish precise and fair-minded choices. Besides, ML calculations depends intensely on the right choice of elements extricated from the dataset, which ended up being troublesome, tedious, and required high computational power. These variables block the presentation of the learning model and produce deadly blunders that can imperil the existences of patients. Conversely, Ismaeel [8] contended that standard measurable methods, the work insight and the instinct of clinical specialists prompted unwanted inclinations and blunders while recognizing gambles related to the sickness. With the significant flood of electronic wellbeing information, clinical

specialists are confronting difficulties to distinguish infections precisely at a beginning phase. Hence, progressed computational techniques, for example, ML calculations were acquainted with find significant examples and concealed data from information, which can be utilized for basic navigation. In outcome, the weight on the clinical staff diminished, while the endurance pace of patients was enhanced [3], [8].

1.2 Aim

The aim of this review is to test the proposed speculation that directed ML calculations can further develop medical services by the exact and early location of infections. In this review, we examine concentrates on that use more than one directed ML model for every sickness acknowledgment issue. This approach delivers more exhaustiveness and accuracy in light of the fact that the assessment of the presentation of a solitary calculation over different review settings prompts inclination which creates uncertain outcomes. The examination of ML models will be directed on couple of sicknesses situated on a heart, kidney, breast, and chronic disease. For the recognition of the disease, various strategies will be assessed like KNN, NB, DT, CNN, SVM, and LR. Toward the finish of this writing, the best performing ML models in regard of every infection will be closed

1.3 Machine Learning

AI is enthusiastically recommendable and broadly utilized in different fields like Security, finance, medical services and so forth. AI: the exemplary definition is - A PC program is said to gain for a fact E concerning a few class of errands T and execution measure P, on the off chance that its presentation at assignments in T, as estimated by P, improves with experience E . A kind of AI gives frameworks the capacity to gain and grow naturally for a fact without being unequivocally customized [10]. Machine inclining is the means by which PCs perceive examples and settle on choices without being expressly modified. With ML, rather than programming a PC bit by bit, we can program a PC to learn very much like we learn, through experimentation, and heaps of training. 2.1. Job of ML in Prediction Machine Learning gains for a fact. "Experience" is right here "bunches of information". It can take any sort of information - pictures, video, sound or text and start to perceive design in that information. With the assistance of AI method, machines figure out how to deal with the information all the more productively as we can't remove the significant data from information simply by survey it, yet AI can do [11]. Its motivation is to gain from information [12]. A ton of explores have been done on how machine advance without help from anyone else [13] . When it figures out how to perceive designs in information, it can likewise figure out how to make expectations in light of those examples. AI can possibly significantly further develop forecast which is much of the time utilized related to enormous informational indexes. It comprises of such countless productive calculations, structures and applications to accomplish more prominent rightness of forecast.

1.3.1 Stages to apply Machine Learning on data

(I) **Information Gathering**: Either the information is composed on paper, archived in text records and calculation sheets, or put away in a data set framework, there is generally a prerequisite to deal with it in an electronic organization so it very well may be reasonable for investigation. This progression is exceptionally basic on the grounds that the quality and measure of information that we get, straightforwardly impacts the nature of ML project. This information will act as the learning material utilized by a calculation to create significant information.

(ii) **Information Preparation**: Once the information is assembled, then, at that point, it is stacked into a reasonable spot and set it up for use in ML preparing.

(iii) **Picking a Model**: Data researchers and analysts have previously made such countless models. These models are appropriate for successions (viz text or music), picture information, text-based information and mathematical information. A fitting calculation will be chosen and the information as a model be addressed.

(iv) **Preparing**: By the time model is picked and information has been arranged for investigation, then it will be utilized to consistently work on the capacity of a model to precisely foresee the outcomes.

(v) **Assessment:** As ML models lead to a one-sided answer for the issue of learning, it is thusly critical to decide how well the calculation has gained from its insight. This should be possible by testing the model against that information which has not

been utilized for preparing. This assists us with perceiving how the model will neutralize information that has not yet seen. Preparing - assessment is typically separated in the scope of 70%-30% or 80%-20%.

(vi)   **Working On Model Performance**: Once the assessment is done, it is conceivable that we need to additionally work on the preparation. This should be possible by tuning a portion of the boundaries.

Consequently to work on model execution, testing is done on various suppositions.

 Forecast: After all the above advances have been finished, Prediction is the last advance to   accomplish something valuable by sending the model. ML utilizes information to respond to questions, so expectation is the stage where a few inquiries can at last be replied. Hence the worth of ML is acknowledged in the entirely of this work.

1.3.2 Machine Learning Types

(I**) Supervised Learning**: It is that sort of ML calculation which utilizes a referred to dataset likewise alluded to as preparing dataset to frame orders or forecasts [14]. This dataset incorporates marked information that comprise of info information and reaction values. For instance, arrangements are furnished along with each issue.

(ii)   **Unsupervised Learning**: It is that sort of ML calculation which is utilized to draw surmising's from informational indexes comprising of information without marked responses [15]. For example, when children begin removing choices from their own comprehension.

(iii) **Reinforcement Learning**: It is that sort of Machine Learning which gains from its own experience [16]. For instance, assuming another circumstance comes up, youngster will make moves all alone, from the previous experience, yet parent can figure out if the activity is great or not.

Aside from all over three principal calculations, such countless calculations are likewise there that are the piece of these three ML calculation types.

1.3.3 Applications of Machine Learning

(i)   **Speech Recognition (SR):** Machine learning assists the product with adjusting to dynamic discourse designs clients use sayings shoptalk, contractions and to remain adaptable. This is where AI is fundamental. Indeed, even hypothetically, a human group can't show a large number of discourse varieties to the product physically. On the off chance that the framework trains itself, but the assignment turns out to be considerably more sensible.

(ii)   **Image Recognition**: It is the main use of Machine Learning. It is a methodology for recognizing and distinguishing an article, spot, individual or the element in an advanced picture. Different instances of this strategy are - face identification, design acknowledgment, face acknowledgment, OCR and some more. The utilization of ML is to dissect the picture pixel by pixel and concentrate the elements of a picture. Facebook

(iii) **Prediction**: Prediction is the method involved with deciding something in light of past history. It very well may be house cost forecast, climate expectation, infection expectation, traffic forecast, and some more. Each sort of estimate is conceivable with ML approach. There are such countless calculations used to achieve this assignment. These calculations are helpful in anticipating and diagnosing persistent sicknesses in medical care [17].

(iv) **Sentiment Analysis**: The primary undertaking of Sentiment Analysis is to foresee "others' thought process?". For instance, somebody has composed that "the item is great", then to figure out the specific assessment or thought about that individual from a text that "is it great or not". It is applied on dynamic applications, survey based site. The job of ML is to remove the information from information by utilizing both directed and solo learning.

(v)   **Healthcare Services**: ML strategies acquires an advancement medical care industry. It is broadly utilized in clinical issues to infection forecast and determination, clinical exploration, treatment, backing and arranging. As per scientists, ML assumes a significant part in precise distinguishing proof of infection that assistance to work with clinical specialists so the

nature of clinical consideration can be moved along. The estimation in ML applications are the consequences of clinical analysis, for example, unique clinical pictures, clinical trials viz. blood test, pulse test and so on, presence or nonattendance of different side effects and general data of patient like age, weight and so forth. Based on these aftereffects of estimation, specialists slender down on the sickness causing the patient.

(vi)   **Video Surveillance**: Machine learning can assist with creating complex calculations for video acknowledgment at first utilizing human oversight. The framework can assist with detecting any dubious items, human figures, obscure vehicles and so forth. Before long it will be feasible to envision a video observation framework that capacities altogether without human oversight.

(vii) **Email Analysis**: Machine calculation can dissect and contrast authentic messages and spam and decide contrasts even in situations where people would effortlessly commit an error.

(viii) **News Classification**: As how much data is developing dramatically, individual client need apparatuses that would characterize and sort the data according to their advantage and decision. Henceforth ML calculations can be helpful to go through huge number of articles in numerous dialects and select the ones that are applicable to client interests and propensities.

(ix)   **Cyber Security**: ML calculations can promptly identify digital protection dangers. The framework will perceived the danger examined comparative cases and take measure to get the site or application. It permits organizations to be fully informed regarding pernicious practices and foresee security issues before they even come up.

(x)   **Social Media Services**: Social media like Facebook and numerous other are utilizing methods of Machine Learning to ceaselessly screen our exercises and in view of the exercises observed give us such countless alluring highlights like - idea to remark, whom to talk, companions ideas and so on.

(xi)   **Recommendation System**: This is one of the development uses of ML. Many web crawlers like Google and web based shopping sites are involving this element wherein comparative kind of sites, items and administrations are recommender to a client after a buy or search. Different Machine Learning calculations are utilized to carry out suggestion framework.

(xii) **Author Identification**: As we realize that the utilization of Internet is hugely developing, because of which its unlawful use for unseemly objects is a central issue now-a-days. Subsequently, ML calculation helps in creator distinguishing proof so violations can be halted.

(xiii) **Online Customer Service**: To give online client care, sites foster a visit BOT as a delegate to deal with client questions. This should be possible with the assistance of Machine Learning calculations; it examinations client conduct from the information visit. Bot designers can realize which issues to zero in on. When a few many reactions were affirmed, the talk BOTS can advance all alone from everyday cooperation with clients getting better with every exchange.

(xvi) **Language Identification**: Machine Learning is most productive methodology in distinguishing the kind of a Language. Apache Tika, Apache Open NLP are the most well-known programming utilized for language ID.

(xv) **Information Retrieval**: As we realize that information is developing hugely on the web, consequently IR assumes a significant part in large information. It is the strategy for extricating from unstructured information significant data. ML utilizes client propensities and interests from dissecting search measurements and gives the outcome. Presently, evaluating calculations will not depend on Meta labels and catchphrases yet rather will examine the setting of the page. Google RankBrain is the extraordinary illustration of this thought.

(xvi) **Age/Gender Identification**: Machine Learning is progressed to the point that it can assist with recognizing age as well as orientation of an individual by utilizing one of the calculation is SVM classifier. This application is a lot of valuable in scientific assignment.

(xvii)   **Robot Control**: To deal with the highlights of robot, helicopter, robot and so on, AI calculations are broadly utilized in robot control framework.

(xviii)   **Virtual Personal Assistant**: ML calculations are likewise utilized in Personal Assistant. It can break down private information, process voice demands, robotize day to day task and can adjust the adjustment of client needs. For instance, Alexa by Amazon utilize all gathered information to further develop its example acknowledgment abilities and have the option to address new requirements based on experience.   **(xix) Self-driving vehicles**: It is among the most well-known Machine Learning application. ML calculations are utilized to prepare the vehicle models with the goal that it can identify the items and individuals while driving.

(xx) Traffic Prediction: This is among those use of ML which is utilized in our day to day routine. This application is useful in foreseeing the traffic information all the more precisely. It is conceivable due to ML, which stores every one of the continuous information and utilizations it to figure number of vehicles and their speed out and about and afterward perform traffic forecast.

II. RELATED WORK

Harshit Jindal et.al(2021)[18] Step by step the instances of heart sicknesses are expanding at a quick rate and it's vital what's more, disturbing to foresee any such infections ahead of time. This analysis is troublesome assignment for example it ought to be performed unequivocally and effectively. The exploration paper basically centers around which patient is more liable to have a coronary disease in view of different clinical qualities. We arranged a coronary disease expectation framework to anticipate whether the patient is probably going to be determined to have a coronary disease or not utilizing the clinical history of the patient. We utilized various calculations of AI, for example, calculated relapse and KNN to anticipate and group the patient with coronary disease. A very helpful approach was utilized to direct the way that the model can be utilized to work on the exactness of forecast of Coronary failure in any person. The strength of the proposed model hushed up fulfilling and was capable to anticipate proof of having a coronary disease in a specific person by utilizing KNN and Logistic Relapse which showed a decent precision in contrast with the recently utilized classifier, for example, credulous bayes and so forth. So a calm critical measure of strain has been lift off by utilizing the given model in tracking down the likelihood of the classifier to accurately and precisely recognize the coronary disease. The Given coronary disease forecast framework improves clinical consideration and decreases the expense. This venture gives us huge information that can assist us with foreseeing the patients with coronary disease It is executed on the .pynb design.

Vijeta Sharma et.al(2020) [19] According to the new concentrate by WHO, heart related infections are expanding. 17.9 million individuals pass on each year because of this. With developing populace, it gets further hard to analyse and begin treatment at beginning phase. Be that as it may, because of the new headway in innovation, Machine Learning procedures have sped up the wellbeing area by different investigates. Hence, the target of this paper is to assemble a ML model for coronary disease expectation in view of the connected boundaries. We have involved a benchmark dataset of UCI Heart infection forecast for this exploration work, which comprise of 14 unique boundaries connected with Heart Disease. AI calculations, for example, Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree have been utilized for the advancement of model. In our exploration we have additionally attempted to find the connections between the various traits accessible in the dataset with the assistance of standard Machine Learning techniques and afterward utilizing them proficiently in the expectation of chances of Heart sickness. Result shows that contrasted with other ML methods, Random Forest gives more precision quicker than expected for the forecast. This model can be useful to the clinical specialists at their facility as choice emotionally supportive network.

Shweta Agarwal et.al(2021) [20] The best way to defeat with the mortality because of persistent sicknesses is to foresee it prior so the disease avoidance should be possible. Such model is a Patient's need where Machine Learning is energetically recommendable. The primary goal is to gather all distributed articles connected with infection forecast and close about the inclusion of exploration done so far. For leading this study, we zeroed in on distributed research from 2017 up to the present time, survey the examination done on different AI calculations utilized for the effective expectation of disease. It has been seen that elements, free factor choice and mix of various calculations assumes a significant part in working on the precision as well as execution of a sickness expectation framework, and it is feasible to analyse individuals in view of symptoms. In this paper, we talked about various ML methods and their exactness that different scientists used to analyse persistent disease.

Changgyun et.al(2019) [21] The point of this study was to anticipate constant disease in individual patients utilizing a person repetitive brain organization (Char-RNN), which is a profound learning model that treats information in each class as a word when a huge piece of its feedback values is absent. A benefit of Burn RNN is that it requires no extra ascription strategy since it verifiably construes missing qualities considering the relationship with neighboring data of interest. We applied Char-RNN to order cases in the Korea National Health and Nutrition Examination Survey (KNHANES) VI as ordinary status and five ongoing infections: hypertension, stroke, angina pectoris, myocardial localized necrosis, furthermore, diabetes mellitus. We additionally utilized a multi-facet perceptron network for a similar assignment for correlation. The outcomes show higher precision for Char-RNN than for the customary multi-facet perceptron model. Scorch RNN showed noteworthy execution in tracking down patients with hypertension also, stroke. The current review used the KNHANES VI information to show a down to earth way to deal with foreseeing and overseeing constant sicknesses with somewhat noticed data.

Rayan Alanazi(2022) [22] These days, people face different infections because of the ongoing ecological condition and their living habits. The identification and forecast of such disease at their previous stages are much significant, to forestall its limit. It is challenging for specialists to physically distinguish the disease precisely more often than not. The objective of this paper is to distinguish and foresee the patients with more normal constant ailments. This could be accomplished by utilizing a state of the art AI method to guarantee that this arrangement dependably distinguishes people with persistent sicknesses. The forecast of disease is likewise a difficult errand. Consequently, information mining assumes a basic part in disease forecast. 'e proposed framework offers a wide sickness guess in light of patient's side effects by utilizing the AI calculations, for example, convolutional brain organization (CNN) for programmed highlight extraction also, sickness expectation and K-closest neighbor (KNN) for distance computation to observe the specific match in the informational index and the last infection expectation result. An assortment of infection side effects has been performed for the readiness of the information set alongside the individual's living propensities, and subtleties connected with specialist discussions are considered in this broad disease forecast. At long last, a similar investigation of the proposed framework with different calculations, for example, Naïve Bayes, choice tree, and strategic relapse has been exhibited in this paper.

Mohammad Monirujjaman Khan et.al(2022) [23] Quite possibly the most common and driving reasons for malignant growth in lady is bosom disease. It has now turned into a continuous medical condition, what's more, its commonness has as of late expanded. The most straightforward way to deal with managing bosom malignant growth discoveries is to remember them from the beginning. Early location of bosom malignant growth is worked with by PC supported identification and determination (CAD) advances, which can help individuals carry on with longer lives. The significant objective of this work is to exploit ongoing improvements in CAD frameworks and related strategies. In 2011, the United States detailed that one out of each and every eight ladies was determined to have disease. Bosom malignant growth starts because of unusual cell division in the bosom, which prompts either harmless or threatening malignant growth development. Therefore, early location of bosom malignant growth is basic, and with compelling treatment, many lives can be saved. This research covers the discoveries also, examinations of different AI models for distinguishing bosom cancer. The Wisconsin Breast Cancer Diagnostic (WBCD) dataset was utilized to foster the strategy. Regardless of its little size, the dataset gives a few fascinating information. The data was investigated and put to use in various AI models. For forecast, irregular woods, calculated relapse, choice tree, what's more, K-closest neighbor were used. Whenever the outcomes are thought about, the calculated relapse model is found to offer the best results. Strategic relapse accomplishes 98% exactness, which is superior to the past technique revealed.

Vinoth S et.al(2020) [24] Applications of AI (ML) have been expanding generally in different fields like suggestion, shortcoming ID and infection forecast. In ML, various calculations were accessible and used in disease forecast like coronary disease expectation, malignant growth expectation and different structures infection expectation. In our proposed work, bosom disease expectation utilizing ML has been executed. At first bosom disease pictures have been taken as information, pre-processing steps will be done to eliminate boisterous and unessential information from picture. Then, at that point, 2D middle channel is a nonlinear activity frequently utilized in picture handling to lessen "salt and pepper" clamor. To expand difference of picture contrast-restricted versatile histogram evening out is utilized. Division has been carried out and GLCM include extraction is conveyed in light of this data characterization is executed. For exact grouping Artificial Neural Network (ANN) is utilized to anticipate regardless of whether the patient is impacted by bosom malignant growth. Contrasted with other existing technique our strategy predicts brings about precise way.

F. M. Javed Mehedi Shamrat et.al(2020) [25] The focal part of this review is to assess the different Machine learning classifier's exhibition for the expectation of bosom malignant growth disease. In this work, we have involved six directed characterization strategies for the grouping of bosom malignant growth infection. For instance, SVM, NB, KNN, RF, DT, and LR utilized for the early forecast of bosom malignant growth. In this way, we assessed bosom malignant growth dataset through responsiveness, particularity, f1 measure, and absolute exactness. The expectation execution of bosom malignant growth investigation shows that SVM got the highest presentation with the greatest amount of order exactness of 97.07%. While, NB and RF have accomplished the second most elevated exactness by expectation. Our discoveries can assist with diminishing the presence of bosom malignant growth sickness through fostering an AI based prescient framework for early forecast.

Gokul S et.al(2013) [26] This paper proposes the use of a Fully Complex-Valued Radial Basis Function organization (FCRBF), Meta-Cognitive Fully Complex-Valued Radial Premise Function organization (McFCRBF) and Extreme Learning Machine (ELM) for the expectation of Parkinson's infection. With the assistance of Unified Parkinson's Disease Rating Scale (UPDRS), the seriousness of the Parkinson's sickness is anticipated and for untreated patients, the UPDRS scale traverses the reach (0-176). The FC-RBF network utilizes a completely intricate esteemed initiation work sect, which maps c^n to c. The execution of the complex RBF network relies upon the quantity of neurons and introduction of organization boundaries. The execution of the self-administrative learning system in the FC-RBF network results in Mc-FCRBF organization. It has two parts: a mental part and a meta-mental part. The meta-mental part chooses how to learn, what to realize and when to learn in light of the information gained by the FC-RBF organization. Outrageous learning instrument utilizes sigmoid actuation work furthermore, it works with quick speed. In ELM organization, the genuine esteemed sources of info and targets are applied to the organization. The outcome shows that the Mc-FCRBF network has great expectation precision than ELM and FC-RBF network.

Wu Wang et.al (2017) [27] Accurately distinguishing Parkinson's sickness (PD) at a beginning phase is positively imperative for slowing down its progress and giving patients the chance of getting to disease altering treatment. Towards this end, the premotor stage in PD ought to be painstakingly checked. An inventive deep learning strategy is acquainted with early uncover regardless of whether an individual is impacted with PD in view of premotor highlights. In particular, to reveal PD at a beginning phase, a few pointers have been viewed as in this review, including Rapid Eye Movement and olfactory misfortune, cerebrospinal liquid information, and dopaminergic imaging markers. A correlation between the proposed profound learning model and twelve AI furthermore, outfit learning strategies in light of moderately little information including 183 solid people and 401 early PD patients shows the prevalent identification execution of the planned model, which accomplishes the most noteworthy exactness, 96.45% all things considered. Other than recognizing the PD, we likewise give the component significance on the PD discovery process in light of the Boosting technique.

III. WRITING REVIEW

A. Common Disease

Dahiwade et al. [9] proposed a ML based system that predicts common disease. The symptoms dataset was imported from the UCI ML depository, where it contained symptoms of many common disease. The system used CNN and KNN as classification techniques to achieve multiple disease prediction. Moreover, the proposed solution was supplemented with more information that concerned the living habits of the tested patient, which proved to be helpful in understanding the level of risk attached to the predicted disease. Dahiwade et al. [9] compared the results between KNN and CNN algorithm in terms of processing time and accuracy. The accuracy and processing time of CNN were 84.5% and 11.1 seconds, respectively. The statistics proved that KNN algorithm is under performing compared to CNN algorithm. In light of this study, the findings of Chen et al. [28] also agreed that CNN outperformed typical supervised algorithms such as KNN, NB, and DT. The authors concluded that the proposed model scored higher in terms of accuracy, which is explained by the capability of the model to detect complex nonlinear relationships in the feature space. Moreover, CNN detects features with high importance that renders better description of the disease, which enables it to accurately predict disease with high complexity [9], [28]. This conclusion is well supported and backed with empirical observations and statistical arguments. Nonetheless, the presented models lacked details, for instance, Neural Networks parameters such as network size, architecture type, learning rate and back propagation algorithm, etc. In addition, the analysis of the performances is only evaluated in terms of accuracy, which debunks the validity of the presented findings [9]. Moreover, the authors did not take into consideration the bias problem that is faced by the tested algorithms [9], [28]. In illustration, the incorporation of more feature variables could immensely ameliorate the performance metrics of under performed algorithms [29].

B. Kidney Disease

Serek et al. [30] arranged a relative investigation of classifiers execution for Chronic Kidney sickness (CKD) discovery utilizing The Kidney Function Test (KFT) dataset. In this review, the classifiers utilized are KNN, NB, and RF classifier; their presentation is analysed as far as F-measure, accuracy, and precision. According to investigation, RF scored better in expressions of F-measure and exactness, while NB yielded better accuracy. Regarding this review, Vijayarani [31] expected to recognize kidney disease utilizing SVM and NB. The classifiers were utilized to recognize four sorts of kidney disease in particular Acute Nephritic Syndrome, Acute Renal Failure, Chronic Glomerulonephritis, and CKD. Also, the examination was centered around deciding the better performing arrangement calculation in light of the precision and execution time. From the outcomes, SVM impressively accomplished higher precision than NB, which makes it the better performing calculation. Notwithstanding, NB ordered information with least execution time. Other a few observational investigations likewise centered around finding CKD; Charleonnan et al. [32] and Kotturu et al. [33] inferred that the SVM classifier is the most satisfactory for kidney infections since it manages semiorganized and unstructured information. Such adaptability permitted SVM to deal with bigger highlights spaces, which brought about gaining high exactness while recognizing complex kidney disease. Albeit upheld by discoveries, the end is debilitated by earlier idea that different hyper-boundaries were not tested while assessing the exhibitions of ML calculations. Rayan Alanazi [22] reasoned that the CNN and KNN classifier have the most accuracy rate (95%) contrast with Naïve Bayes(NB), Decision Tree(DT), Logical Regression(LR) showed in figure 1.
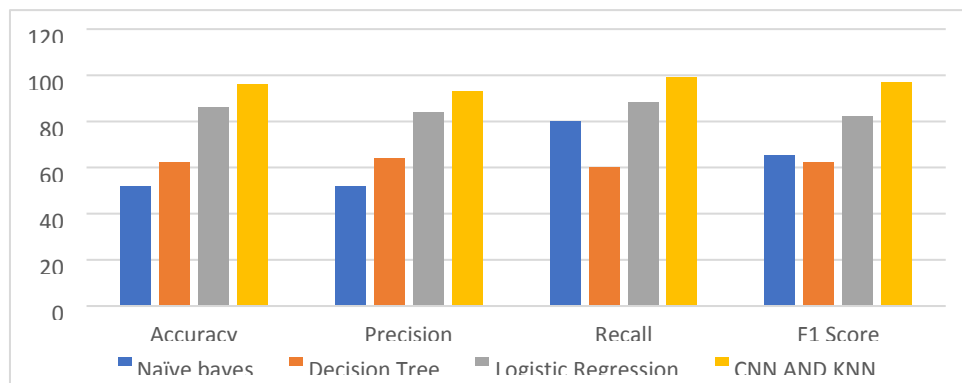

Figure 1: Comparison of other performance evaluation metrics of Logistic and other algorithms.

C. Heart Disease

Marimuthu et al. [34] meant to foresee heart disease utilizing directed ML strategies. The creators organized the properties of information as orientation, age, chest torment, orientation, target and incline [34]. The applied ML calculations that were sent are DT, KNN, LR and NB. According to examination, the LR calculation gave a high exactness of 86.89%, which considered to be the best contrasted with the other referenced calculations. In 2018, Dwivedi [35] endeavored to add more accuracy to the expectation of heart disease by representing extra boundaries, for example, Resting pulse, Serum Cholesterol in mg/dl, and Maximum Heart Rate accomplished. The utilized dataset was imported from the UCI ML research facility; it was contained with 120 examples that were coronary disease positive, and 150 examples that were coronary disease negative. Dwivedi endeavored to assess the exhibition of Artificial Neural Networks (ANN), SVM, KNN, NB, LR and Classification Tree. At the apparatus of ten times cross approval, the outcomes showed that LR has the most noteworthy order exactness and awareness, which shows high reliability at distinguishing heart disease [35]. This end is fortified by the discoveries of Harshit Jindal [18] that the Logistic Regression and K-Nearest Neighbor (KNN) were outflanked (88.5%) from different strategies like SVC, DT, KNN, RF, LR. The investigations succeeded in leading a broad examination on the ML models. For example, different hyper-boundaries were tried at every ML calculation to meet to the most ideal exactness and accuracy values. In spite of that benefit, the little size of the imported datasets imperatives the gaining models from focusing on disease with higher exactness and accuracy. The following

'Figure 2', 'Figure 3', 'Figure 4 'and 'Figure 5' shows a plot of the number of patients that are been segregated and predicted by the classifier depending upon the age group, resting blood pressure, sex, chest pain:
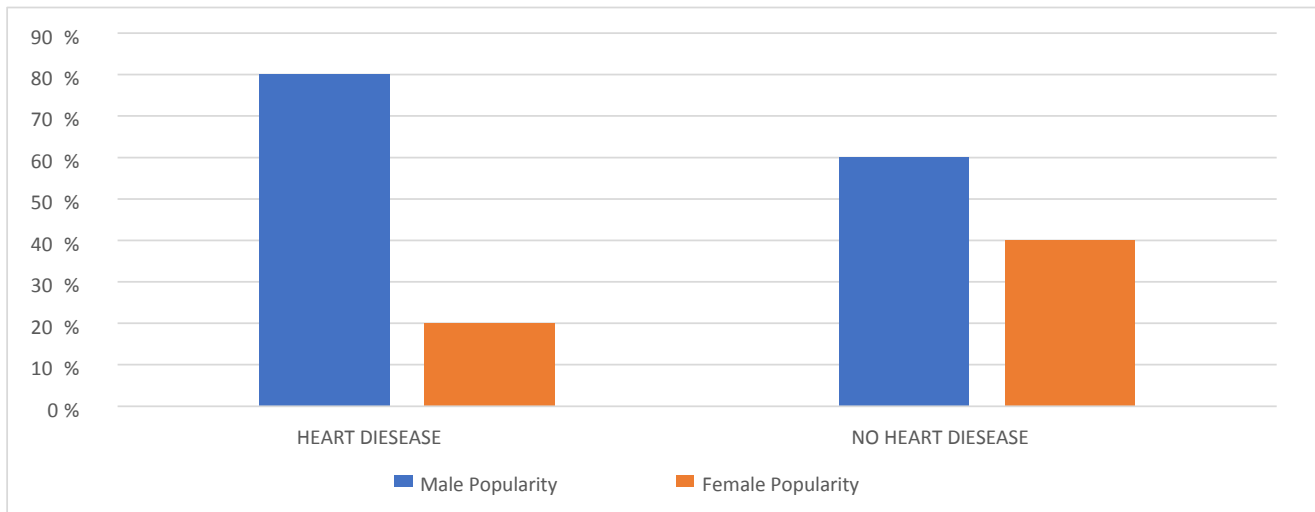
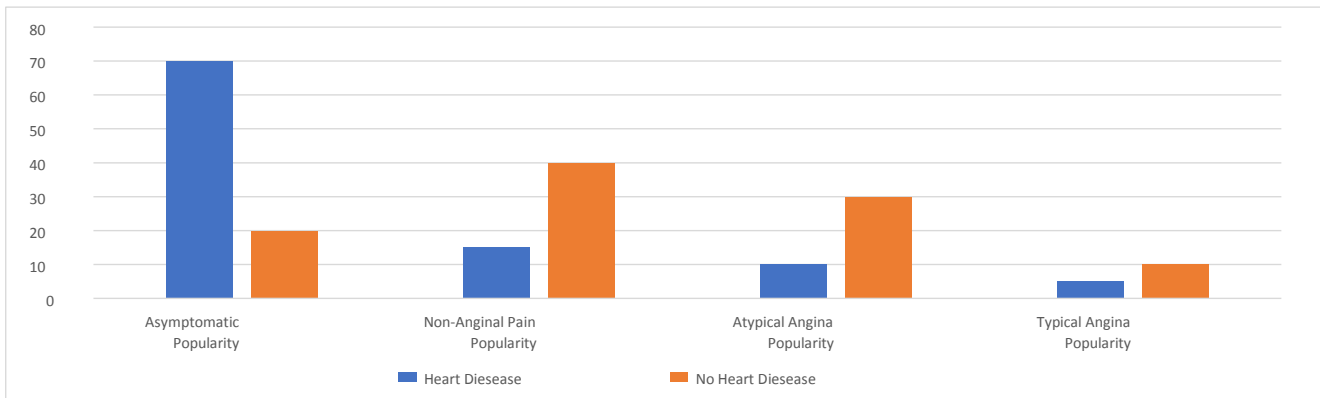**Figure 2:** Shows The Risk of Heart attack on the basis of age

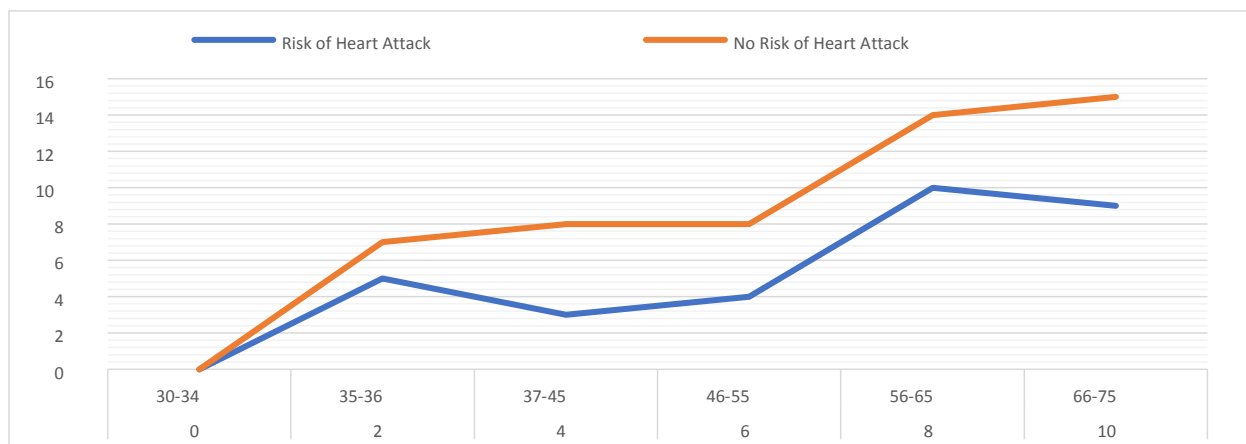**Figure 3:** Shows the Risk of Heart Attack on the basis of the Popularity

Figure 4: Total Number of Patients Having or Not Having Heart Disease Based on Popularities

D. Breast Disease

Shubair [36] endeavored for the recognition of bosom malignant growth utilizing ML calculations, to be specific RF, Bayesian Networks and SVM. The scientists acquired the Wisconsin unique bosom disease dataset from the UCI Repository and used it for looking at the learning models as far as key boundaries like exactness, review, accuracy, and area of ROC diagram. The classifiers were tried utilizing K-overlay approval technique, where the picked worth of K is equivalent to 10 [36]. The re-enactment results have demonstrated that SVM succeeded concerning review, exactness, and accuracy. Notwithstanding, RF had a higher likelihood in the right order of the growth, which was suggested by the ROC chart. Conversely, Yao [37] explored different avenues regarding different information mining strategies including RF and SVM to decide the most appropriate calculation for bosom malignant growth forecast. Per results, the classification rate, sensitivity, and specificity of Random Forest calculation were 96.27%, 96.78%, and 94.57%, individually, while SVM scored a accuracy worth of 95.85%, an sensitivity of 95.95%, and a specificity of 95.53%. However Manirujjiaman et al. [23] pointed that the Logical Regression have the more more accuracy of 98% contrast with different calculations, for example, Random Forest, Decision Tree and K-Nearest Neighbor (Showed in following figure 5, figure 6, figure7 and figure 8). As per Yao [37], discarding portions of information diminishes the nature of pictures, and in this manner the presentation of the ML calculation is ruined
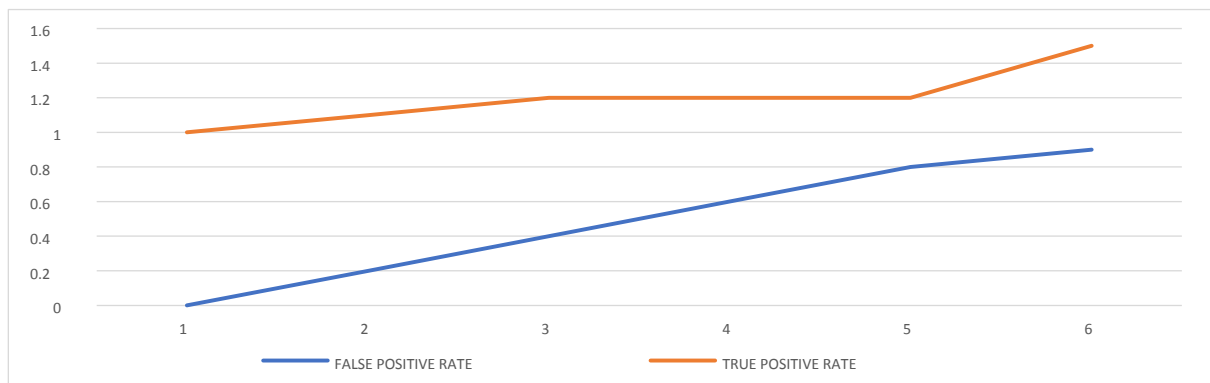

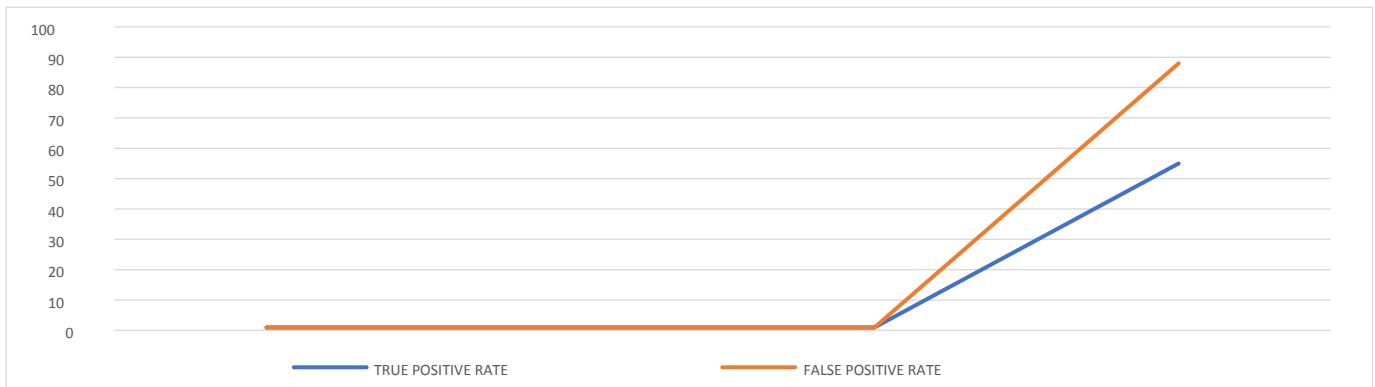**Figure 5:** Random Forest Model Classification Report

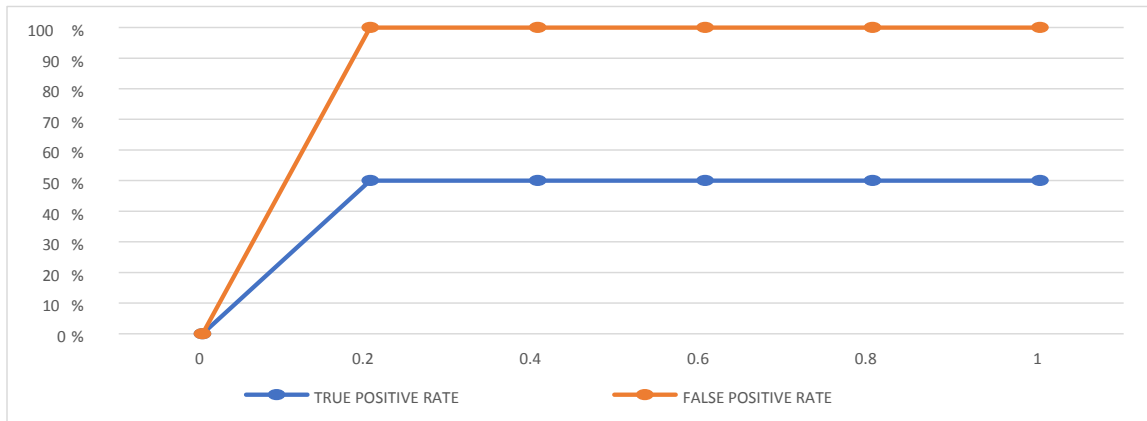**Figure 6:** Decision Tree Model Classification Report

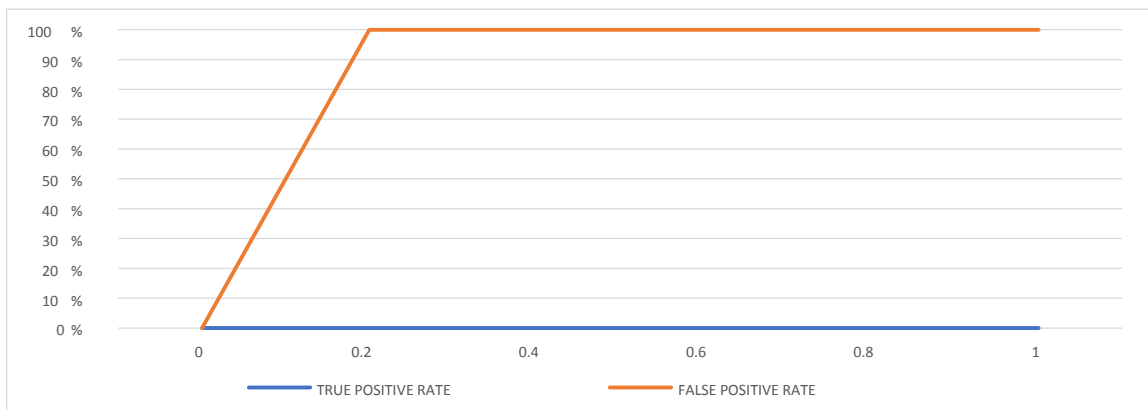**Figure 7:** KNN Model Classification Report

Figure 8: Logistic Regression Model Classifier Report

E. Parkinson's disease

Chen et al. [38] brought a feasible evaluation framework utilising Fuzzy k-Nearest Neighbor (FKNN) for the belief of Parkinson`s infection (PD) . The overview zeroed in on searching on the proposed SVMprimarily based totally and the FKNN-primarily based totally approaches. The Principal Component Analysis (PCA) became used to gather the maximum separated highlights for the improvement of a super FKNN model. The dataset became taken from the UCI vault, and it recorded numerous biomedical voice estimation going from 31 individuals, 24 with PD. The exploratory discoveries have tested that the FKNN method favourably accomplishes over the SVM philosophy with reference to awareness, exactness, and 24 with PD. In line of this overview, Behroozi [39] deliberate to endorse some other grouping shape to research PD, which became stepped forward via way of means of a channel primarily based totally spotlight preference calculation that improved the characterization precision as much as 15%. The characterization of the device became portrayed via way of means of making use of unfastened classifiers for each subset of the dataset to symbolize the deficiency of tremendous data. The picked classifiers have been KNN, SVM, Discriminant Analysis and NB. The effects confirmed that SVM carried out the maximum improved in all of the exhibition measurements. Also, Eskidere [40] targeted on following the motion of PD via way of means of inspecting the presentation of SVM with distinctive classifiers like Least Square Support Vector (LS-SVM), General Regression Neural Network (GRNN) and Multi-layer Perceptron Neural Network(MLPNN). The discoveries tested that LS-SVM is the maximum noteworthy appearing model. This quit is bolstered via way of means of the RIcha Mathur [50] that's the KNN calculation with ANN offers stepped forward final results for foreseeing Parkinson's disease. As indicated via way of means of

Lavesson [41], distinctive ML calculations are meant to improve numerous execution measurements (e.g., Neural Networks improves squared mistake aleven though KNN and SVM streamline exactness). Moreover, the creators are specially extraordinary at featuring systems with subtleties. For instance, SVMs boundaries, for example, the piece and the regularization esteem have been illustrated internal and out. In any case, ML fashions have been now no longer aligned previous to assessing the exhibitions. Caruana contends that [42] adjustment drastically enhancements the grouping of now no longer many mastering fashions specially NB, SVM, and RF.

Ⅳ PERFORMANCE METRICES

Execution Metrics Performance measurements [43] is one of the way that analyst use in proposed Machine Learning Algorithm to check the efficiency, execution and effectiveness. For Classification Problem

Performance metric which are utilized to assess the characterization issue forecasts are -

(i)   Confusion Matrix: Used to decide an exactness of a calculation, it is a two layered "Genuine" and "Anticipated" table. By utilizing the amount of the corner to corner of a network, the all-out quantities of accurately anticipated values are determined and all the others are considered as incorrect[44]-[46].

The terms referenced in disarray framework are made sense of as -

*   True Positive: Shows the presence of disease in a patient when it is really sure.

*   True Negative: Shows the shortfall of sickness in a patient when it is really negative.

*   False Positive: Shows the presence of sickness in a patient when it is really negative.

*   False Negative: Shows the shortfall of disease in a patient when it is really certain.

Table: 1 Confusion Matrix

	Predicted	
Actual	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

(ii)   Accuracy: This is the total number of forecasts that are precise over any remaining expectations. With the assistance of an equation, the Accuracy can be determined as-

$$Accuracy = \frac{TN+TP}{TN+FP+FN+TP}(1)$$

(iii)   Precision: It is chiefly utilized in Information Retrieval, it lets us about know extent of model really sure. With the assistance of an equation, the Precision can be determined as-

$$Precision = \frac{TP}{FP+TP}(2)$$

(iv)   Recall/True Positive Rate/Sensitivity: It educates us concerning the quantity of up-sides returned by the model. With the assistance of an equation, the Recall can be determined as-

$$Recall = \frac{TP}{FN+TP} \quad (3)$$

(v)   Specificity/True Negative Rate: It enlightens us regarding the quantity of negatives that the model returns. With the assistance of a recipe, the Specificity can be determined as-

$$Specificity = \frac{TN}{FP+TN} \quad (4)$$

(vi)   F1 Score: It is a solitary score that addresses both review and accuracy by giving their consonant mean.

For F1 score, the best worth would be 1 and most terrible would be 0.

$$F1 \ Score = 2 * \frac{Recall * Precision}{Precision + Recall}(5)$$

(vii) AUC Curve: AUC implies Area under ROC Curve. It is an exhibition metric used to characterize the precision of a calculation.

(viii) ROC: ROC is a likelihood bend while AUC ascertains the distinctness that characterizes a classifier's presentation. The plot of a TP rate is drawn against a FP rate in a graph displayed in figure 4. Higher the worth of AUC, better will be the model. The nature of ROC Curve is determined in view of the accompanying four constraints[47]-[49].

- On the off chance that the AUC goes from 0.9 to 1, the nature of the test is phenomenal.

- On the off chance that the AUC goes from 0.8 to 0.9, the nature of the test is Good.

- On the off chance that the AUC goes from 0.7 to 0.8, the nature of the test is Fair.

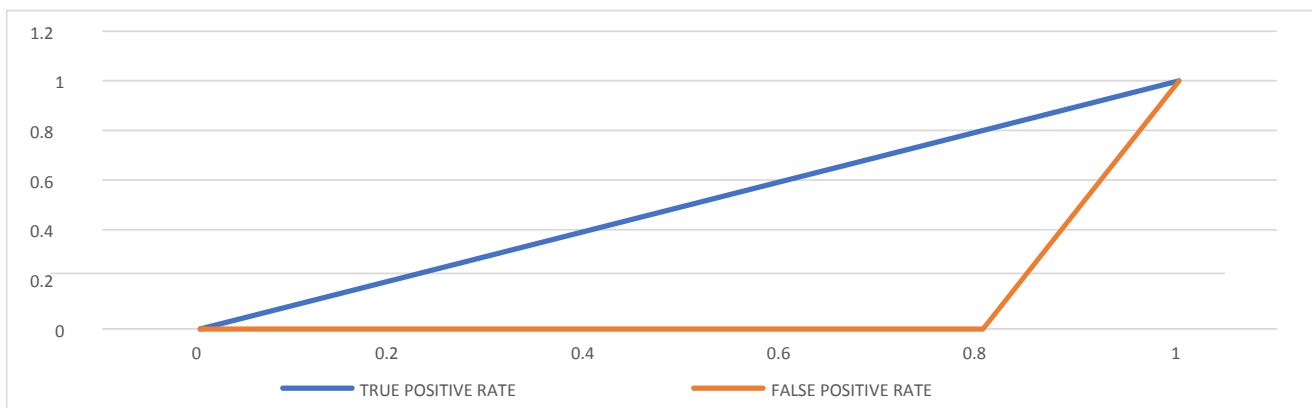- On the off chance that the AUC goes from 0.6 to 0.7, the nature of the test is Poor.

Figure 9 : True positive rate against False positive rate

(ⅶ) Mean Square Error (MSE): One of the favored measurements for relapse errands is MSE or Mean Squared Error. The squared contrast between the objective worth and the anticipated worth can be determined as a normal.

(ⅷ) Root Mean Square Error (RMSE): RMSE is the most normally utilized relapse task metric and is the square foundation of the typical square distinction between the model's objective worth and the anticipated worth. It very well may be determined as -

Ⅴ CONCLUSION

The utilization of various ML calculations empowered the early recognition of numerous disease like heart, kidney, bosom, and mind infections. All through the writing, KNN, CNN and LR calculations were the most generally utilized at forecast, while precision was the most utilized presentation metric. The CNN model ended up being the most sufficient at foreseeing normal disease. Moreover, SVM model showed prevalence in precision at most times for kidney infections and PD in light of its dependability in dealing with high-layered, semi-organized and unstructured information. For Breast malignant growth expectation, LR showed greater prevalence in the likelihood of right characterization of the disease in light of its capacity to scale well for huge datasets and its weakness to stay away from overfitting. At long last, the LR calculation ended up being the most solid in foreseeing heart sickness.

In future work, the production of more complicated ML calculations is genuinely necessary to expand the productivity of disease expectation. Moreover, learning models ought to be adjusted all the more frequently after the preparation stage for possibly a superior execution. In addition, datasets ought to be developed various socioeconomics to keep away from overfitting and increment the exactness of the sent models. At long last, more applicable element choice techniques ought to be utilized to upgrade the presentation of the learning models.

REFERENCES

[1] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275–1278.

[2] Y. Hasija, N. Garg, and S. Sourav, "Automated detection of dermatological disorders through image processing and machine learning," in 2017 International Conference on Intelligent Sustainable Systems (ICISS), 2017, pp. 1047–1051.

[3] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1– 16, 2019.

[4] R. Katarya and P. Srinivas, "Predicting heart disease at early stages using machine learning: A survey," in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302–305.

[5] P. S. Kohli and S. Arora, "Application of machine learning in disease prediction," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1–4.

[6] M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai, and R. Mishra, "A proposed model for lifestyle disease prediction using support vector machine," in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1–6.

[7] F. Q. Yuan, "Critical issues of applying machine learning to condition monitoring for failure diagnosis," in 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2016, pp. 1903–1907.

[8] S. Ismaeel, A. Miri, and D. Chourishi, "Using the extreme learning machine (elm) technique for heart disease diagnosis," in 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015), 2015, pp. 1–3.

[9] D. Dahiwade, G. Patle, and E. Meshram, "Designing disease prediction model using machine learning approach," Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, no. Iccmc, pp. 1211–1215, 2019.

[10] C. G. Raji and S. S. Vinod Chandra, "Long-Term Forecasting the Survival in Liver Transplantation Using Multilayer Perceptron Networks," IEEE Trans. Syst. Man, Cybern. Syst., vol. 47, no. 8, pp. 2318– 2329, 2017, doi: 10.1109/TSMC.2017.2661996.

[11] W. Richert and L. P. Coelho, Building Machine Learning Systems with Python.

[12] Ö. Çelik, "A Research on Machine Learning Methods and Its Applications," J. Educ. Technol. Online Learn., vol. 1, no. 3, pp. 25–40, 2018, doi: 10.31681/jetol.457046.

[13] M. Metcalf, "A first encounter with f90," ACM SIGPLAN Fortran Forum, vol. 11, no. 1, pp. 24–32, 1992, doi: 10.1145/134304.134306.

[14] R. Konieczny and R. Idczak, "Mössbauer study of Fe-Re alloys prepared by mechanical alloying," Hyperfine Interact., vol. 237, no. 1, pp. 1–8, 2016, doi: 10.1007/s10751-016-1232-6.

[15] B. Rao, "Machine Learning Algorithms: A Review," Int. J. Comput. Sci. Inf. Technol., vol. 7, no. 3, pp. 1174–1179, 2016.

[16] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," J. Artif. Intell. Res., vol. 4, pp. 237–285, 1996, doi: 10.1613/jair.301.

[17] G. Winter, "Machine learning in healthcare harvesting of results that a consultant," vol. 25, no. 2, pp. 100–101, 2019.

[18] Harshit Jindal, Sarthak Agarwal, Rishabh Khera, Rachana Jain, Preeti Nagrath, " Heart Disease Prediction Using Machine Learning Techniques", 2021, IOP Conference Series: Materials Science and Engineering

[19] Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta, " Heart Disease Prediction Using Machine Learning Techniques" 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)

[20] Shweta Agarwal, Dr.Chander Prabha, Dr.Meenu Gupta, "Chronic Disease Using Machine Learning Techniques- A Review", Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 3495 - 3511 Received 15 December 2020; Accepted 05 January 2021.

[21] Changgyun Kim , Youngdoo Son and Sekyoung Youm, "Chronic Disease Prediction Using CharacterRecurrent Neural Network in the Presence of Missing Information", 2019, MDPI.

[22] Rayan Alanazi," Identification and Prediction of Chronic Diseases Using Machine Learning Approaches", Hindawi Journal of Healthcare Engineering, Volume 2022, Article ID 2826127, 9 pages https://doi.org/10.1155/2022/2826127

[23] Mohammad Monirujjaman Khan et.al, " Machine Learning Based Comparative Analysis for Breast Cancer Prediction", Hindawi Journal of Healthcare Engineering, Volume 2022, Article ID 4365855, 15 pages, https://doi.org/10.1155/2022/4365855

[24] Vinoth S and Valarmathi P ," Accurate Breast Cancer Prediction Using Machine Learning Techniques", International Journal of Recent Technology and Engineering(IJRTE) ISSN: 2277-3878 (Online), Volume-8, Issue-6, March-2020

[25] F. M. Javed Mehedi Shamrat et.al, " An Analysis On Breast Disease Prediction Using Machine Learning Approaches", International of Scientific & Technology, ISSN:2277-8616 Research Volume 9, Issue-02, February 2020.

[26] Wu Wang et.al, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning", Office of Sponsored Research (OSR), 2019

[27] Gokul S et.al," Parkinson Disease Prediction Using Machine Learning Approaches", 2013 Fifth International Conference on Advanced Computing (ICoAC)

[28] S. Jadhav, R. Kasar, N. Lade, M. Patil, and S. Kolte, "Disease Prediction by Machine Learning from Healthcare Communities," International Journal of Scientific Research in Science and Technology, pp. 29–35, 2019.

[29] R. Saravanan and P. Sujatha, "A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification," in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 945– 949.

[30] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), 2018, pp. 1–4.

[31] V. S and D. S, "Data Mining Classification Algorithms for Kidney Disease Prediction," International Journal on Cybernetics & Informatics, vol. 4, no. 4, pp. 13–25, 2015.

[32] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," 2016 Management and Innovation Technology International Conference, MITiCON 2016, pp. MIT80– MIT83, 2017.

[33] P. Kotturu, V. V. Sasank, G. Supriya, C. S. Manoj, and M. V. Maheshwarredy, "Prediction of chronic kidney disease using machine learning techniques," International Journal of Advanced Science and Technology, vol. 28, no. 16, pp. 1436–1443, 2019.

[34] M. Marimuthu, M. Abinaya, K. S., K. Madhankumar, and V. Pavithra, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach," International Journal of Computer Applications, vol. 181, no. 18, pp. 20–25, 2018.

[35] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing and Applications, vol. 29, no. 10, pp. 685–693, 2018.

[36] P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for breast cancer prediction," Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, pp. 796–801, 2020.

[37] D. Yao, J. Yang, and X. Zhan, "A novel method for disease prediction: Hybrid of random forest and multivariate adaptive regression splines," Journal of Computers (Finland), vol. 8, no. 1, pp. 170–177, 2013.

[38] H. L. Chen, C. C. Huang, X. G. Yu, X. Xu, X. Sun, G. Wang, and S. J. Wang, "An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach," Expert Systems with Applications, vol. 40, no. 1, pp. 263–271, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2012.07.014

[39] M. Behroozi and A. Sami, "A multiple-classifier framework for Parkinson's disease detection based on various vocal tests," International Journal of Telemedicine and Applications, vol. 2016, 2016.

[40] O. Eskidere, F. Ertas¸, and C. Hanilc¸i, "A comparison of regression¨ methods for remote tracking of Parkinson's disease progression," Expert Systems with Applications, vol. 39, no. 5, pp. 5523–5528, 2012.

[41] N. Lavesson, Evaluation and Analysis of Supervised Learning Algorithms and Classifiers, 2006.

[42] R. Caruana and A. Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics," Proceedings of the 23rd international conference on Machine Learning, pp. 161–168, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.3232

[43] N. M. J. Kumari and K. K. V. Krishna, "Prognosis of Diseases Using Machine Learning Algorithms: A Survey," Proc. 2018 Int. Conf. Curr. Trends Towar. Converging Technol. ICCTCT 2018, pp. 1–9, 2018, doi: 10.1109/ICCTCT.2018.8550902.

[44] M. Kaur, H. K. Gianey, D. Singh, and M. Sabharwal, "Multi-objective differential evolution based random forest for e-health applications," Mod. Phys. Lett. B, vol. 33, no. 5, 2019, doi: 10.1142/S0217984919500222.

[45] J. P. Singh and R. S. Bali, "A hybrid backbone based clustering algorithm for vehicular ad-hoc networks," in Procedia Computer Science, 2015, vol. 46, pp. 1005–1013, doi: 10.1016/j.procs.2015.01.011.

[46] N. Mittal, U. Singh, and B. S. Sohi, "A Novel Energy Efficient Stable Clustering Approach for Wireless Sensor Networks," Wirel. Pers. Commun., vol. 95, no. 3, pp. 2947–2971, 2017, doi: 10.1007/s11277017-3973-1.

[47] P. Gairola, S. P. Gairola, V. Kumar, K. Singh, and S. K. Dhawan, "Barium ferrite and graphite integrated with polyaniline as effective shield against electromagnetic interference," Synth. Met., vol. 221, pp. 326– 331, 2016, doi: 10.1016/j.synthmet.2016.09.023.

[48] G. Sharma, S. Sharma, and S. Gujral, "A Novel Way of Assessing Software Bug Severity Using Dictionary of Critical Terms," in Procedia Computer Science, 2015, vol. 70, pp. 632–639, doi: 10.1016/j.procs.2015.10.059.

[49] M. K. Gupta et al., "Parametric optimization and process capability analysis for machining of nickelbased superalloy," Int. J. Adv. Manuf. Technol., vol. 102, no. 9–12, pp. 3995–4009, 2019, doi: 10.1007/s00170-019-03453-3.

[50] Richa Mathur et.al, " Parkinson Disease Prediction Using Machine Learning Algorithms", 2018, Part of the Advances in Intelligent Systems and Computing book series (AISC,volume 841)