Medical Record-Based Health Estimation with Organ-Focused Probability Analysis

Anuj Yadav[a], Tejas Mahajan[a], Dr Pramod Kumar Maurya[a*],

a: School of Computer Science and Engineering, Vellore Institute of Technology, Vellore

Abstract: The healthcare industry is constantly developing, and there is a need for precise and qualified methods for medical treatment. Machine learning algorithms likely offer faster processing of medical data as compared to the analysis done by the doctors manually. The main idea of the project is to develop and implement a ml-based system that will analyze the medical report for kidney, liver, and blood related disease in the industry of healthcare. The system makes use of the natural language processes techniques for extracting the data, which is relevant from the medical reports, and then it is preprocessed and with the help of machine learning algorithm it is classified. The health status of the patient is presented in the concise dashboard made using sensible decisions by the healthcare professionals with the help of comparison of the models. The most proficient model among these will be used to predict the presence of a particular disease in the patient. This streamlines healthcare operations, saves time for analysis of medical reports, and overall efficiency is improved.

Keywords- Liver Disease, Kidney Disease, Heart Disease, Machine Learning Algorithms, Extra Trees Classifier

I. Introduction

The background of this project lies in the intersection between healthcare technology, machine learning and natural language processing. The project is used for harnessing the power of machine learning and natural language for automating the analysis of medical reports which helps in reducing the burden on the doctors and improves the accuracy in the diagnosis of the disease. The main idea is to provide faster processing of the medical data in comparison with the manual analysis by the doctors.

Chronic kidney disease is a condition in which the kidney does not work properly, and the filtration of blood does not take place as usual. [7] The main work of the kidney is to get rid of the waste material and extra water from the blood. For balancing the minerals and salts in our body, filtration is very important. CKD means the waste that is built in the body. CKD is an internationally public health problem affecting the population of the world around 5-10%. Chronic kidney disease is a significant non-communicable disease that contributes to morbidity and mortality on a global scale through cardiovascular diseases attributable to impaired kidney function [1]. One of the most important organs of the human body is the kidney. It is accountable for the purification of blood. A damaged kidney makes a person sad. CKD Disease will increase the risk of life and cause other health diseases like heart disease, diabetes, high blood pressure, etc. CKD can also lead to kidney transplants. The signs and symptoms of the CKD disease in the initial states are foamy urine, itchy or dry skin, feeling tired, loss of appetite, weight loss without trying to lose weight. When a person has the advanced stages of CKD then it will have trouble concentrating, shortness of breath, vomiting take place, trouble sleeping, cramps in the muscles. There are five stages of chronic kidney disease (CKD): -

Stage 1: eGFR >= 90
Stage 2: eGFR between 60 &89
Stage 3: eGFR between 30 & 59
Stage 4: eGFR between 15 & 29
Stage 5: eGFR <15

Heart disease describes a scale of conditions that affect our heart. Today, cardiovascular diseases are the main cause of death worldwide with 17.9 million deaths yearly, according to World Health Organization reports

**Email Id: Anuj Yadav(anujyadavnnl03@gmail.com , Tejas Sharad Mahajan(tejasmahajan101@gmail.com) ,
Pramod Kumar Maurya(pramodkumar.maurya@vit.ac.in)**

the diagnosis of heart disease is a challenging task, which can offer predictions about the heart condition of a patient so that further treatment can be made efficiently. [8] There are a range of heart diseases like blood vessel disease, irregular heartbeat, congenital heart defects, heart valve disease, and heart muscle disease. The symptoms of heart disease in blood vessels are chest pain, shortness of breath, pain in the neck, jaw, and in back. The symptoms of irregular heartbeats are dizziness, fluttering in the chest, fainting, lightheadedness. The symptoms caused by the congenital heart defects are pale gray or blue skin or maybe lips, swelling in the area of the belly, around the eyes and in the legs., less weight gain. [14] The major risk factors of heart disease are smoking, unhealthy diet taking, high blood pressure, high cholesterol, diabetes, stress. We can prevent heart disease by exercising 30 min a day, reducing stress, controlling high blood pressure, diabetes, by taking at least 9 hours of sleep every day, and by consuming a diet which is low in salt and fat.

Liver is located at the upper part of the gastrointestinal tract of the body, and its range in women is 1200-1400 g and for men it is 1400-1800g. [10] It is the largest gland in the body. The main work of the liver is for digestion, metabolism, release of toxins, and immunization. Liver diseases are broken down on the basis of their aetiology and effect on the liver. Cirrhosis is also the main disease of the liver. Liver diseases are associated with alcohol or hepatitis. The symptoms of liver disease are swelling and pain in the belly, swelling in the legs and ankles, dark urine, pale stone, vomiting, tiredness, loss of appetite, eyes color becoming whiter, and skin color changes to yellow. [2] We can prevent liver disease by not consuming alcohol, by getting vaccinated, by being careful while taking medicines, by having distance from the people who have body fluids and hepatitis, by maintaining a healthy weight. Liver diseases affect billions worldwide, posing a notable burden on the healthcare industry and individual lives. Liver diseases consist of a scale of conditions ranging from fatty liver disease to cirrhosis, posing significant challenges to healthcare systems worldwide.

II. LITERATURE REVIEW

In the research by Md. A. Islam et al. [1], there are 850 million people worldwide who are likely to have renal disease due to various causes. At least 2.4 million individuals every year pass away from kidney-related diseases, per the World Kidney Day 2019 study. The study of data is from St. Paulo’s Hospital, 441 (25.67%) instances are end-stage renal disease stage or stage five, 399 (23.22%) are at a severe stage or stage four, 354 (20.61%) stage three, 248 (14.44%) stage two, and 276 (16.07%) have no chronic kidney disease. The class distribution of binary class is 1442 (83.93%) CKD (stage 1 to 5) and 276 (16.07%) not CKD. The binary-class distribution is imbalanced. Data oversampling. The value of the minority class and the value of the majority class have been balanced using resampling techniques. After using the resampling procedure, the binary class dataset's total size increased to 2888. Feature selection used are Unsupervised Forward Selection (UFS) and Recursive Feature Elimination RFECV/RFE. A total of 18 models trained. 8 features - 99.8% accuracy - RF with RFECV.

M.F.Rabbi et al. used the [2] the dataset from UCI ILPD. Random oversampling is used to overcome imbalanced class distribution. Z-score outlier is used to detect outliers using robust scaling in inter-quartiles. (replacing the mean and standard deviation with the median and normalized IQR) or simply median over mean. The best accuracy is when ad boosting is used over classical ML approach, with extremely randomized trees - 91.48%.

[3] In this research, 7 classifiers were used. However, Logistic and KNN did not give suitable results and it was why they were not used in SMOTE. As per the result, it is concluded that SMOTE is the best technique for balancing a dataset. It is noted that SMOTE gave better results with selected features by LASSO

regression as compared to without SMOTE on LASSO regression model. LSVM achieved the highest accuracy in all experiments as compared to other classifiers algorithms.

Balogh EP et al. [4] studied the definition of diagnostic error used by the committee is presented, together with information on the measuring strategy used by the committee and the epidemiology of diagnostic mistakes, including both cognitive and system errors. Newman Toker - a conceptual model, where when the diagnosis that a patient receives is incorrect or no attempt to provide a diagnosis label, effort is taken to avoid label failures that occur. Usually occurs when failure develops a prompt and correct explanation of the patient's health issue. Even overutilization and over diagnosis causes increased healthcare costs, overtreatment, mostly in places with high populations, however it is a challenge to the healthcare sector for scope of improvement. Health insurance claims also help in making accurate data for processing in various fields, given the diagnosed issue was timely and data integrity is maintained and collected via bills. The rising use of health information technology has made measurement of process and quality more accurate.

The author N.Bora et al. in [5] uses Two datasets for the analysis: one with 303 records and 14 attributes from the UCI database, and another of 1190 records and 12 attributes from Kaggle. The features were normalized by using standard scalers. An 80/20 dataset split was done, and the ML algorithms were used to predict accuracy in each case. A combined dataset was also created, which is considered a valuable system since it contains data from two different regions and sources, and hence involves different forms of data. However, the relative accuracy may decrease due to dataset complexity, Random Forest gives the best testing accuracy of 93.31% and training accuracy of 100%.

The research was done by ILYAS et al. [6] they used the dataset from UCI CKD and tool used is WEKA data mining tool. Preprocessing based on Glomerular Filtration Rate, based on age, sex, race and serum creatinine, are selected from the dataset to be given as input in GFR calculation. J48(decision tree algorithm - ID3) and Random Forest classifier along with k-FOLD cross validation on data with value of 15. Both algorithms used give 96% accuracy.

[7] This study starts USES 24 parameters WITH class attribute, and ends up WITH 30 % of them as sub set for predicting Chronic Kidney Disease. 4 ML based classifiers have been examined within a supervised learning setting, achieving highest performance is AUC 0.995, sensitivity 0.9897, and specificity 1. The experimental procedure ended that advances in machine learning, with assist of predictive analytics, represent a promising setting which to see intelligent solutions, which in turn prove the ability of predication of the kidney disease.

Nagavelli U et al. proposed [8] Dataset is used from ECG with timestamp for time series, in form of an excel sheet, translated into weights and stored. Records were formed by clusters. 18 features were summarized.
Proposed system is a web application with user-friendly features and pages .4 levels on each of the 36 channels are applied by db-4 (Daubechies 4) DWT. Algorithm: used are XG boost, SVM, naive bayes Output is accuracy parameter is high in the XG Boost and low in the NB with weighted approach.

In Paper by Lei N, Zhang X et al. [9] The paper is unique for its dataset preparation, using free searches and NLP methods to extract relevant information. Dataset Prep- PubMed, EMBASE, Cochrane Central Register of Controlled Trials, the Chinese Biomedicine Literature Database, Chinese National Knowledge Infrastructure, Wan fang Database, and VIP Database were used both free-text terms and Medical Subject Headings (MeSH), and data was converted using Rev Man 5.2. For the data both the bivariate model and the hierarchical summary receiver operating characteristic (HSROC) method for data synthesis. A total of 184,052 articles were collected and 188 taken for the final dataset after processing. Algorithms - (LR), (SVM) and (RF) algorithm and ANN. Quality assessment tool used - QUADAS Results - RF with max AUC of 0.87 and ANN with 0.933.

The authors Rakshith D B et al. On topic [10] Liver Disease Prediction System using Machine Learning Techniques Dataset - UCI ILPD, with 167 individuals who do not have liver disease and 415 patients who do, and 10 variables including age, gender, total proteins, albumin, A/G ratio, SGPT, SGOT, and Alkphos. Tool - PyQt 5 for large classes and functions is a python-based GUI. Algorithms used - in Image using SK Learn python library. Output- 100% with SVM

[11] This paper by Braun LT et al. talks about clinical reasoning improvement in the educational sector, analysis hence of medical students is talked about .4th and 5th year medical students from two medical schools in Munich Reasons for errors mentioned are with diagnostic skills (24%), inadequate knowledge base (16%), premature closure (10%), faulty context generation (15%). Ten technical tests. The order in which the students graded the history, physical, and technical assessments was up to them too. Participants were misdiagnosed 304 out of 704 times. Issues mentioned - The study examined 8 distinct examples, and it is still unknown whether a sample of more cases will reveal the same mistake categories. Plus, these were internal cases, and they don't know whether this standard can be applied to medical discipline The incidence of diagnostic error in medicine: Arthur Elstein, a cognitive psychologist - concluded the diagnosis is wrong 10–15% of the time, Autopsy studies identify major diagnostic discrepancies in 10–20% of cases. According to estimates, 2-4% of the time, test results are misleadingly incorrect. For instance, a systematic examination of over 8000 ER patients discovered that 9% had a delayed diagnosis of stroke in a few countries. It is found that patients are both willing and capable of participating effectively in identifying errors in their care. Correct data collection technique, avoiding disease malpractice, insurance claim databases need to be rectified.

[12] The experimental results show that Multiclass Decision Forest algorithm gives more result the other classification algorithms and gives 99.17% accuracy.

Sen.R et al. [13] proposes RPN that shares features with a detection network for object detection. The RPN generates region proposals, which uses Fast R-CNN for object detection. The RPN predicts object bounds and scores of each position and is for generating high-quality region proposals. Sharing the convolutional features, the RPN, Fast R-CNN are merged into a single network that has state-of-the-art object detection. The system has a rate of 5fps on a GPU of the VGG-16 model. R-CNN and RPN methods used for in the 1st-place entries in tracks of the ILSVRC and COCO 2015 challenge.

Sivakannan Subramani et al. done the paper on cardiovascular disease prediction. [14] Dataset preparation - Search strategy on MEDLINE, Embase, and Scopus databases was applied, keywords used to search for studies of ML algorithms and coronary heart disease, stroke, heart failure, and cardiac arrhythmias and finally imported in Covidence, an online systematic review tool. Final extraction was carried on - authors, year of publication, study name, test types, testing indications, analytic models, number of patients, endpoints and performance measures ((AUC, sensitivity, specificity, positive cases, negative cases, true positives, false positives, true negatives, and false negatives)) Output - 45 cohorts reported a total of 116,227 individuals. 10 cohorts used CNN algorithms, 7 cohorts used SVM, 13 cohorts used boosting algorithms, 9 cohorts used custom-built algorithms, and 2 cohorts used RF. The prediction of CAD was AUC 0.88, sensitivity with 0.86, specificity with 0.70 of boosting algorithms & pooled of AUC 0.93.

[15] Incomplete data is filled ,6 ml algorithms (LR, random forest, SVM, k-nearest neighbor, naive Bayes

classifier and feed forward neural network) are used to establish models. machine learning models, random forest achieved the best with 99.75% diagnosis accuracy. Misjudging by analysis generated by the established models, we have done an integrated model that combines LR and random forest with help of perceptron, which could get an average accuracy of 99. 83% after 10 times of evaluation.

III. METHODOLOGY

The overall project can be divided into 3 phases:
The blood report contains the input as in image format. There are steps for generating the dataset for health prediction. Steps to be followed is below:

Scanning Medical Test Report: In this step, we will scan the medical test report for conversion in digital form. The scanned copies are supposed to be input data.

As shown in Figure 1, OCR Scan Py tesseract Library: After the report is scanned, the OCR (Optical character Recognition) is done with the help of a popular OCR engine known as Py tesseract that provides accurate results for text recognition. OCR is a technique that will recognize the text characters from the scanned images. and then convert them into machine-readable text

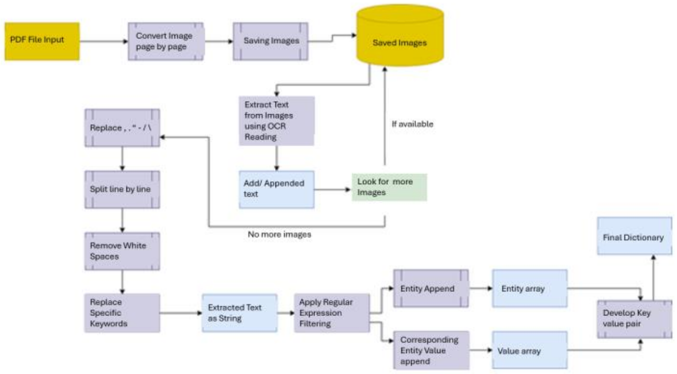

Figure 1: Proposed Text Extraction Flow

Regex Filtering: After that, the OCR Scanned output generated is passed through a series of regex (Regular Expression) filters. Regex is a pattern matching technique used for identifying specific patterns in the text. Regex filters are used for extracting relevant information from the OCR output such as patients' name, age, gender, test type, test result etc.

NLP Processing: NLP is the branch of artificial intelligence which deals with the interaction between computers and human language. The system uses NLP Techniques to extract more meaningful information from the text data i.e. (medical condition information, symptoms, treatment) from the report. as shown in figure 2.

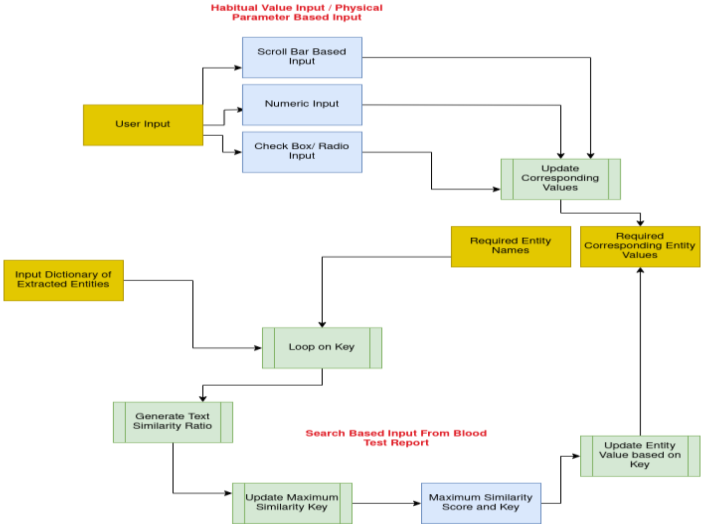
Habitual Value Input / Physical Parameter Based Input
Scroll Bar Based Input
Numeric Input
User Input
Check Box/ Radio Input
Update Corresponding Values
Required Entity Names
Required Corresponding Entity Values
Input Dictionary of Extracted Entities
Loop on Key
Generate Text Similarity Ratio
Search Based Input From Blood Test Report
Update Entity Value based on Key
Update Maximum Similarity Key
Maximum Similarity Score and Key

Figure 2: Proposed Text Extraction Flow

Dataset Generation: Once the information is extracted, it is organized in a structured dataset. The dataset consists of patients' information.

Model Training: As show in figure 3, All the models are trained for better efficiency and the null values are replaced for better accuracy.

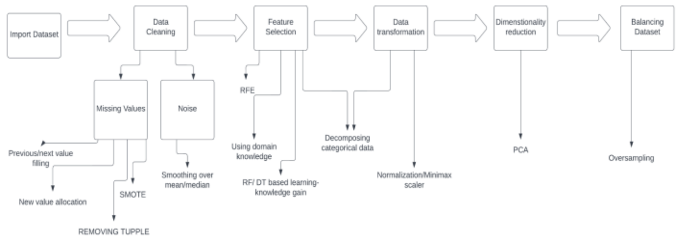

Figure 3: Proposed Flow for Training the ML Models

We have compared multiple models, and the more proficient one is being used for predicting the disease. It will display the result on the streamlite dashboard.

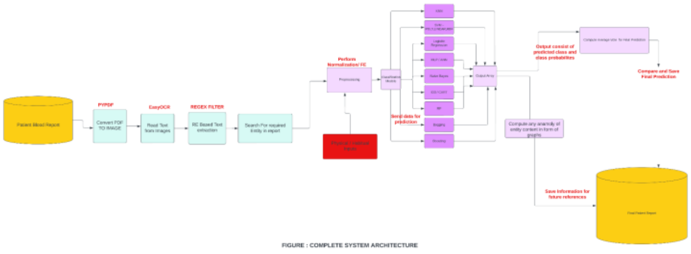
Patient Blood Report
PYPDF
Convert PDF TO IMAGE
EasyOCR
Read Text from images
REGEX FILTER
RE Based Text extraction
Search For required Entity in report
Perform Normalization/ FE
Preprocessing
Physical / Habitual Inputs
Send data for prediction
Output Array
Output consist of predicted class and class probabilities
Compare and Save Final Prediction
Compute any anamoly of entity content in form of graphs
Save Information for future references
Final Patient Report
FIGURE : COMPLETE SYSTEM ARCHITECTURE

Figure 4 Proposed Architecture Model

IV. DATASET DESCRIPTION

Liver:

This data set is made up of 416 liver patient records and 167 non liver patient records. Groups are divided into whether patient has liver disease or not by using selector class label. This data set consists of 441 male patient records and 142 female patient records. Any patient whose age exceeded 89 is listed as being of age "90". There are 10 input features in this dataset and one final output feature defining if the person is healthy or not.

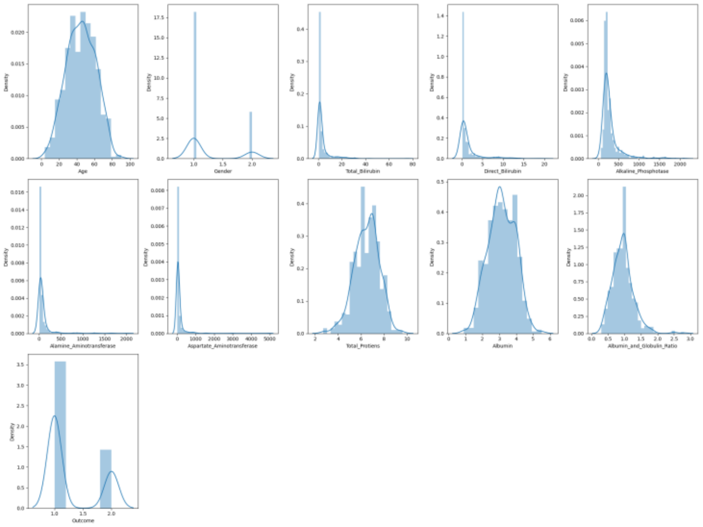

Figure 5: Graph Representation of Attributes of Liver Dataset

In figure 5, The attributes taken in the Liver dataset are age, gender, Total Bilirubin, Direct Bilirubin, Alkaline phosphatase, Alamine aminotransferase, Aspartate aminotransferase, Total proteins, Albumin, Albumin and Globulin Ratio, Outcome and the graphical representation of the dataset.

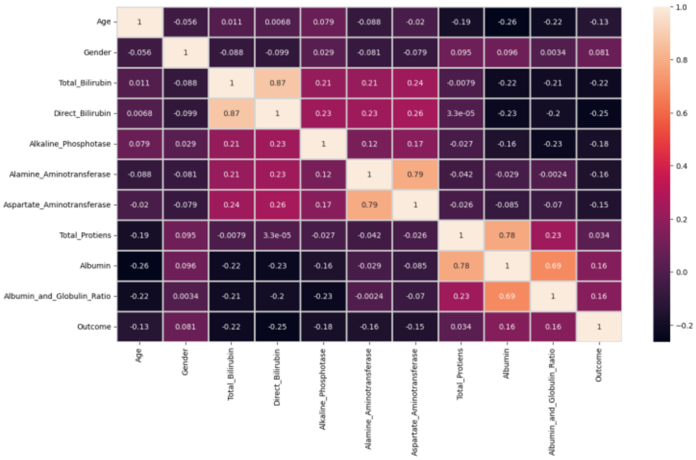

Figure 6: Pearson Correlation Matrix for Liver Failure Dataset

The correlational matrix for Liver is shown in figure 6 between the data set attributes and its value.

Kidney:

There are 25 input features in this dataset and one output feature defining the disease. The Chronic Kidney Dataset comprises 400 instances where 38%

percent of the records are of healthy people and 62% is of people affected by CKD.

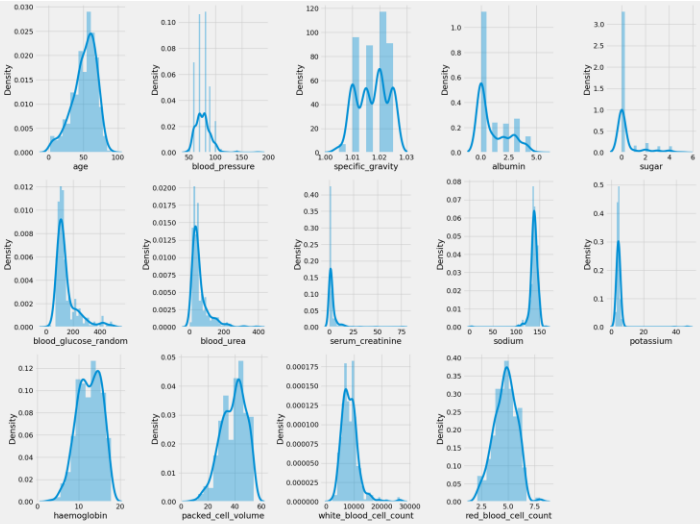

Figure 7: Graph Representation of Attributes of Kidney Dataset

The figure 7 shows the attributes taken in the Kidney dataset are age, blood pressure, specific gravity, albumin, sugar, blood glucose random, blood urea, serum creatinine, sodium, potassium, red blood cell count, white blood cell count, packed cell volume, hemoglobin, and the graphical representation of the dataset.

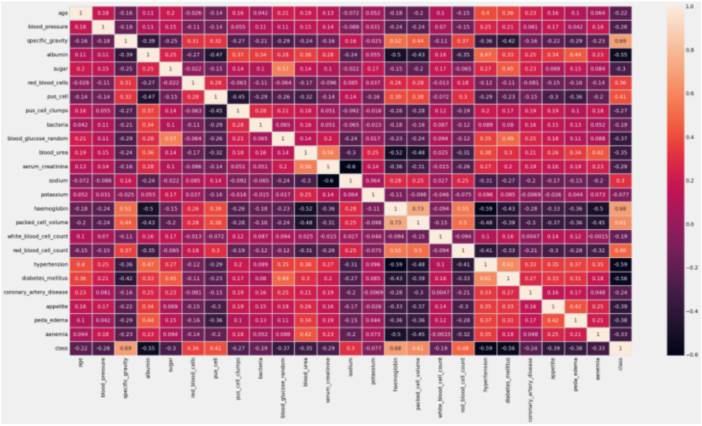

Figure 8: Pearson Correlation Matrix for Kidney Failure Dataset

The correlation matrix for kidney is shown in figure 8 between the data set attributes and its value.

Heart

In our proposed work we have used the dataset containing the medical report records of 303 patients who have probability of heart failure.

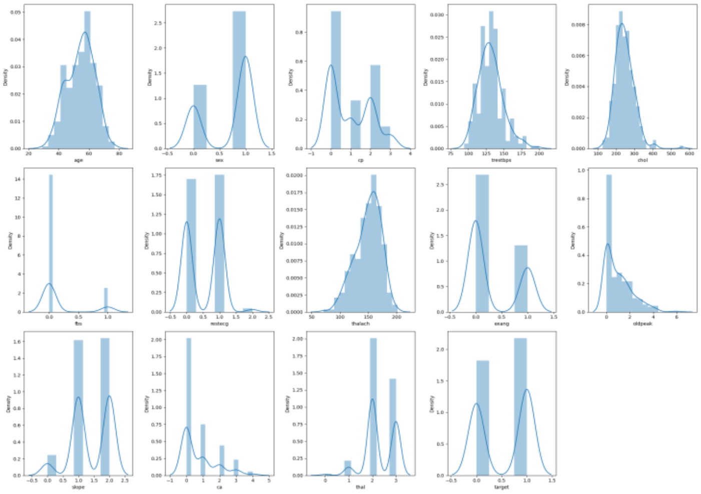

Figure 9: Graph Representation of Attributes of Heart Dataset

The figure 9 shows the attributes taken in the Kidney datasetareage,sex,cp,trestbps,chol,slope,restecg, thalach,examg, oldpeak,ca, Thal, target and the graphical representation of the dataset attributes

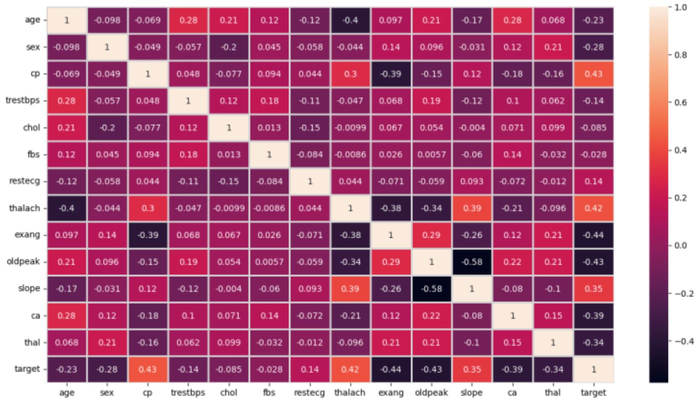

Figure 10: Pearson Correlation Matrix for Heart Failure Dataset

The correlation matrix for heart is shown in figure 10 between the data set attributes and its value.

V. RESULT AND DISCUSSION

We have analyzed the prediction of the three diseases-kidney, liver, and blood-related disease (i.e. heart). We used multiple algorithms for the prediction. After model comparison, the most preferable algorithm was chosen for a more accurate prediction.

Health Board Dashboard: It will display the collected data from the medical reports and the predicted data from algorithms in a clear and concise way with the help of stream lite.

Models Used

Support Vector Machine: (SVM), a machine learning algorithm, is used for regression, linear or nonlinear classification, outlier detection tasks SVMs is used for text, image classification, handwriting identification, face detection, spam detection, and anomaly detection. SVMs can control high-dimensional data as well as nonlinear relationships.

Extra Tree Classifier: This Algorithm creates many decision trees with the sampling for each tree being random, without replacement. Due to this, a dataset for each tree with unique samples is created. A specific number of features are also selected randomly for each tree from the total set of features.

Xg Boost: XG Boost (Extreme Gradient Boosting) is popular due to its speed and performance in structured or tabular data. XG Boost builds trees sequentially, by correcting errors made by the previous trees, thus incrementally improving the model's predictions. To prevent overfitting, it incorporates optimizations like built-in regularization to prevent overfitting and handles missing values and pruning of trees.

Stochastic Gradient Boosting: SGB is a hybrid of the boosting and bagging approaches and uses L-terminal node small trees as base model. A random subsample of the training dataset is selected to fit a tree at each boosting iteration.

Gradient Boosting Classifier: Gradient Boosting joins the weak learners into strong learners, in which new model is trained to lessen the loss function as mean squared error or cross-entropy of the previous model using gradient descent. The gradient of the loss function is calculated w.t.r the predictions of the current ensemble. To minimize this gradient in each iteration, it then trains a new weak model. The ensemble is added with the prediction of the new model and the process is repeated till a stopping criterion is met.

AdaBoost Classifier: AdaBoost short for Adaptive Boosting is a learning that is used in machine learning for classification and regression problems. The main idea of AdaBoost is for training the weak classifier on the training dataset with every increasing classifier giving more weightage to the data points that are misclassified. The final AdaBoost model is decided by combining all the weak classifiers that have been used for training with the weightage given to the models according to their accuracy. The weak model which has the maximum accuracy is given the highest weightage while the model with least accuracy is given a lower weightage.

Random Forest Classifier: Random Forest is an ensemble machine learning algorithm that operates by building multiple decision trees during training and output the average of the predictions from individual trees for regression tasks, or the major vote for classification tasks. It improves by the performance of a single decision tree by minimizing overfitting, thanks to the randomness introduced while the creation of individual trees Random Forest is trained on a random subset of the training data and uses a random subset of features for making splits.

Decision Tree Classifier: A technique that is used for classification and Regression problems, but mainly used for solving Classification problems. It is a tree-structured classifier, internal nodes represent the features of a dataset, branches show the decision rules, and each leaf node represents the output.

Logistic Regression: Logistic regression is a machine learning algorithm that is used for classifying whether instance belongs to a given class or not. Logistic regression is a statistical algorithm that analyzes the relationship between the two factors.

Kidney:

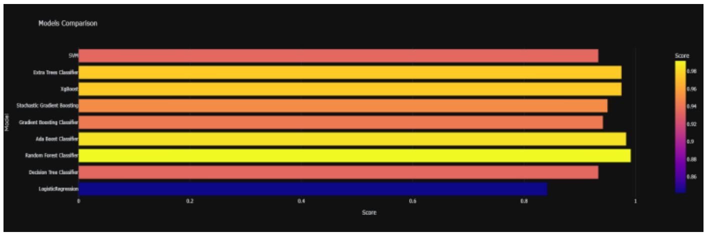

Figure 11: Model Comparison of Algorithms used for Kidney

Liver:

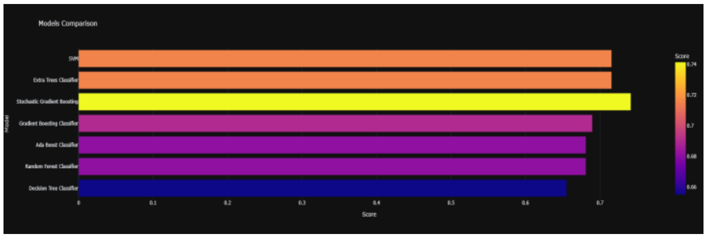

Figure 12: Model Comparison of Algorithms used for Liver

Heart:

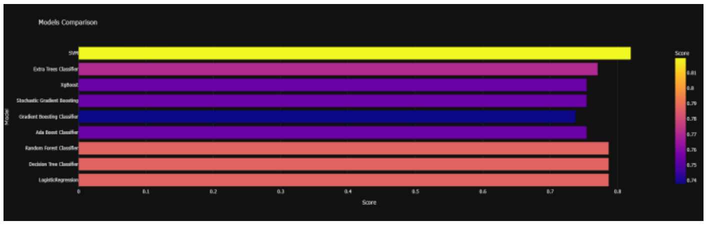

Figure 13: Model Comparison of Algorithms used for Heart

We have compared all the models in Table 1 for all the three diseases and we get the desired result and their accuracy. SVM Model has the maximum accuracy with 98 % in Kidney, 81.9% in Heart and 74% in Liver.

Model	Kidney	Heart	Liver
SVM	98	81.9	74
Extra Tree Classifier	99.1	77.04	68
XgBoost	95.8	75.40	-
Stochastic Gradient Boosting	97.5	73.77	71
Gradient Boosting Classifier	99.6	73.77	70
AdaBoost Classifier	98.33	75.4	77
Random Forest Classifier	96.6	80.32	68
Logistic Regression	85.8	78.68	-
Decision Tree Classifier	87.50	80.32	69

Table:1 Model Comparison for all 3 organ diseases.

VI. CONCLUSION

In Conclusion, the project is based on machine learning that will analyze the medical report for kidney, liver and blood- related diseases. The system extracts the suitable data, classifies it and displays the patient's health information on the dashboard. The most proficient model among these will be used to predict the presence of a particular disease in the patient. With the help of this system, the potential of machine learning will be manifested for improving the patient's health more efficiently. Compared to the traditional methods, this system will be more accurate and show quick diagnosis of the patient.

References

[1] Md. A. Islam, Md. Z. H. Majumder, and Md. A. Hussein, “Chronic kidney disease prediction based on machine learning algorithms,” Journal of Pathology Informatics, p. 100189, Jan. 2023

[2] M.F.Rabbi, S.M.Mahedy Hasan, A.I.Champa, M.Asifzaman, M.K.Hasan, Prediction of Liver Disorders Using Machine Learning Algorihms: A Comparitive Study, 2nd ICAICT, 28-29 November 2020.

[3] P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," in IEEE Access, vol. 9, pp. 17312-17334, 2021

[4] Balogh EP, Miller BT, Ball JR, editors. Overview of Diagnostic Error in Health Care 2015

[5] N.Bora , S.Gutta, A. Hadaegh, Using Machine Learning to predict heart disease, WSEAS Transactions Biology and Biomedicine, VOL. 19, PP. 1–9, JAN. 2022

[6] Ilyas et al, “Chronic Kidney Disease Diagnosis Using Decision Tree Algorithms, BMC, VOL. 22, NO.1, AUG.2021,

[7] Aljaaf, Ahmed J.; Al-Jumeily, Dhiya; Haglan, Hussein M, Alloghani, Mohamed, Baker, Thar, Hussain, Abir J, Mustafina, Jamila, Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics, 1–9, [IEEE 2018 IEEE Congress on Evolutionary Computation (CEC) - Rio de Janeiro (2018.7.8-2018.7.13)]

[8] Nagavelli U, Samanta, D, Chakraborty, P, Machine Learning Technology-Based Heart Disease Detection Models. Journal of Healthcare Engineering, 2022, February 27

[9] Lei N, Zhang X, Wei M, Lao B, Xu X, Zhang M, Chen H, Xu Y, Xia B, Zhang D, Dong C, Fu L, Tang F, Wu Y, Machine learning algorithms' accuracy in predicting kidney disease progression: a systematic review and meta-analysis. BMC Med Inform Decis Mak, 22(1):205, 2022 Aug 1

.

[10] Rakshith D B, Mrigank Srivastava, Ashwani Kumar, Gururaj S P, 2021, Liver Disease Prediction System using Machine Learning Techniques, International Journal of Engineering Research & Technology (IJERT) Volume 10, Issue 06 (June 2021)

[11] Braun LT, Zwaan L, Kiesewetter J, Fischer MR, Schmidmaier R, Diagnostic errors by medical students: results of a prospective qualitative study, BMC Med Educ,17(1):191, 2017 Nov 9

[12] M. P. N. M. Wickramasinghe, D. M. Perera and K. A. D. C. P. Kahandawaarachchi, "Dietary prediction for patients with chronic kidney disease (CKD) by considering blood potassium level using machine learning algorithms,pp. 300-303, 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia, 2017

[13] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[14] Sivakannan Subramani, Neeraj Varshney, M Vijay Anand, Manzoore Elahi M Soudagar, Lamya Ahmed Al-Keridis , Tarun Kumar Upadhyay, Nawaf Alshammari, Mohd Saeed, Kumaran Subramanian, Krishnan Anbarasu, Karunakaran Rohini, Cardiovascular diseases prediction by machine learning incorporation with deep learning, 2023 Apr 17

[15] Qin, Jiongming; Chen, Lin; Liu, Yuhua; Liu, Chuanjun, Feng, Changhao, Chen, Bin, A Machine Learning Methodology for Diagnosing Chronic Kidney Disease, IEEE Access, 1–1, (2019).