** SRI RAMAKRISHNA ENGINEERING COLLEGE**

**BONAFIDE CERTIFICATE**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**PROJECT WORK – MAY 2022**

This is to certify that the project entitled

**HUMAN DISEASE DETECTION USING ENSEMBLE**

**MACHINE LEARNING AND DEEP LEARNING**

is the bonafide record of project work done by

**SHANKARA NARAYANAN K (1801209)**

**SHRUTHHI K (1801215)**

**SUJITH M (1801237)**

of B.E. Computer Science and Engineering during the year 2018-2022.

who carried out the project work under my supervision, certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Head of the Department Project Guide**

Submitted for the Project Viva-Voce Examination held on **\_\_\_\_\_\_\_\_\_\_\_**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Internal Examiner Subject Expert**

**DECLARATION**

We affirm that the project work titled **“<Title in Caps bold>”**being submitted in partial fulfillment for the award of Bachelor of Engineeringis the original work carried out by us. It has not formed the part of any other project work submitted for award of any degree or diploma, either in this or any other University.

---------------- ------------------------------

(Signature of the Candidates)

|  |  |
| --- | --- |
| **SHANKARA NARAYANAN K** | **(1801209)** |
| **SHRUTHHI K** | **(1801215)** |
| **SUJITH M** | **(1801237)** |

I certify that the declaration made above by the candidates is true.

------------------------------

(Signature of the guide)

**T.CHITRA KUMAR,**

**Associate Professor,**

**Department of CSE**

**ACKNOWLEDGEMENT**

We express our gratitude to **Sri.D. LAKSHMINARAYANASWAMY,** Managing Trustee, **Sri. R. SUNDAR,** Joint Managing Trustee, SNR Sons Charitable Trust, Coimbatore for providing excellent facilities to carry out our project.

We express our deepest gratitude to **Dr. N. R. ALAMELU Ph.D.,** Principal, for her valuable guidance and blessings.

We are indebted to our Head of the Department **Dr. A. GRACE SELVARANI Ph.D.,** Professor and Head, Department of Computer Science and Engineering who modeled us both technically and morally for achieving great success in life.

We sincerely thank our Project Coordinator, **Dr.R.MADHUMATHI,** Asst. Professor (Sr.G), Department of Computer Science and Engineering for her great inspiration.

Words are inadequate to offer thanks to our respected guide. We wish to express our sincere thanks to **Dr.P.MATHIYALAGAN,** Associate Professor, Department of Computer Science and Engineering, who gives constant encouragement and support throughout this project work and who makes this project a successful one.

We also thank all the staff members and technicians of our Department for their help in making this project a successful one.

**TABLE OF CONTENTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | | | | **PAGE NO.** |
|  | **ABSTRACT** | | | | **viii** |
|  | **LIST OF FIGURES** | | | | **ix** |
|  | **LIST OF TABLES** | | | | **x** |
|  | **LIST OF ABBREVIATIONS** | | | | **xi** |
| **1** | **INTRODUCTION** | | | | **1** |
|  | 1.1 | Human Disease Prediction using Machine Learning  & Deep Learning | | | 1 |
|  | 1.2 | Problem Statement | | | 4 |
| **2** | **LITERATURE REVIEW** | | | | **7** |
|  | 2.1 |  | | | 7 |
|  | 2.2 |  | | | 7 |
|  | 2.3 |  | | | 8 |
|  | 2.4 |  | | | 9 |
|  | 2.5 |  | | | 9 |
|  | 2.6 |  | | | 10 |
|  | 2.7 |  | | | 10 |
|  | 2.8 |  | | | 11 |
|  | 2.9 |  | | | 12 |
|  | 2.10 |  | | | 12 |
| **3** | **SYSTEM ANALYSIS** | | | | **14** |
|  | 3.1 | Existing System | | | 14 |
|  | 3.2 | Proposed System | | | 15 |
| **4** | **SYSTEM SPECIFICATION** | | | | **16** |
|  | 4.1 | Hardware Requirements | | | 16 |
|  | 4.2 | Software Requirements | | | 16 |
| **5** | **SOFTWARE DESCRIPTION** | | | | **17** |
|  | 5.1 | Python and Flask | | | 17 |
|  | 5.2 | Jupyter Notebook | | | 17 |
| **6** | **PROJECT DESCRIPTION** | | | | **19** |
|  | 6.1 | Problem Definition | | | 19 |
|  | 6.2 | Introduction to Proposed System | | | 19 |
|  |  | 6.2.1 System Architecture | | | 20 |
|  | 6.3 | Module Description | | | 20 |
|  |  | 6.3.1 | Data Pre-Processing | | 21 |
|  |  | 6.3.2 | Importing Modules | | 21 |
|  |  | 6.3.3 | Statistical Analysis Of Data | | 23 |
|  |  | 6.3.4 | Detection using KNN, Regression, XGBoost,  Decision Tree | | 23 |
|  |  | 6.3.5 | Detection using Custom Ensemble Learning | | 23 |
|  |  | 6.3.6 | Website Deployment | | 24 |
| **7** | **SYSTEM IMPLEMENTATION** | | | | **24** |
|  | 7.1 | Disease Detection Using Machine Learning | | | 26 |
|  |  | 7.1.1 | | Detection Using Logistic Regression | 26 |
|  |  | 7.1.2 | | Detection Using Custom Ensemble Learning | 28 |
|  |  | 7.1.3 | | Detection Using Random Forest, Regression | 29 |
|  | 7.2 | Disease Detection Using Deep Learning | | | 30 |
|  |  | 7.2.1 | | Detection Using Convolutional Neural Network | 30 |
|  | 7.3 | Website Deployment | | | 31 |
| **8** | **RESULTS AND DISCUSSION** | | | | **32**  3 |
|  | 8.1 | Experimental Results | | | 32 |
|  |  | 8.1.1 | |  | 33 |
|  |  | 8.1.2 | |  | 33 |
|  | 8.2 | Comparison of Results | | | 34 |
|  | 8.3 |  | | | 37 |
|  | 8.4 |  | | | 38 |
| **9** | **CONCLUSION AND FUTURE ENHANCEMENTS** | | | | **40** |
|  | 9.1 | Conclusion | | | 40 |
|  | 9.2 | Future Enhancements | | | 40 |
|  | **APPENDIX** | | | | **41** |
|  |  | Source Code | | | 41 |
|  |  | Snapshots | | | 46 |
| **10** | **REFERENCES** | | | | **53** |

**ABSTRACT**

Nowadays, India encounters huge varieties of disease and detections of the diseases plays an vital role in medical field. Early-stage detection of diseases is very essential and necessary because of the unhealthy food habits and lifestyle. In this study for predicting disease and to classify the disease based on accuracy which determines the person is healthy or not. To increase the efficiency of predicting the types of disease that the people have and to find the variety of disease which revives in the surroundings, Custom Ensemble Learning using Pipeline, K-Nearest Neighbor (KNN), Logistic Regression, Convolutional Neural Networks (CNN) and Principal Component Analysis is used. The data of the patients were pre-processed., compartmented and analyzed to predict which patient is needs treatment and given priority and which hits the surroundings very most. Ensemble Machine Learning’s model popularity in medicine is used in different approach to predict the accuracy with combiantions of algorithms. To de-escalate the manual processes in medical industry, automating different processes has become very significant. Major Medical records and healthcare advancement have given an opportunity to find out which patients require more significance. Machine Learning were used to pre-process the dataset in order to validate the model. To improve the performance of the model, both Machine learning and Deep Learning is used to predict the accuracy of the disease of the patients, the results are more reliable. The objective of the project is to create a system model to classify patients records whether it is malignant or benign and find out which disease affects the patient’s health and to reduce the cost of the entire processes.

சுருக்கம்

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO** | **NAME OF THE FIGURE** | **PAGE NUMBER** |
| 1.1 |  | 2 |
| 6.1 |  | 21 |
| 6.2 |  | 25 |
| 6.3 |  | 26 |
| 7.1 |  | 27 |
| 7.2 |  | 27 |
| 7.3 |  | 29 |
| 7.4 |  | 30 |
| 7.5 |  | 31 |
| 8.1 |  | 34 |
| 8.2 |  | 35 |
| 8.3 |  | 35 |
| 8.4 |  | 36 |
| 8.5 |  | 37 |
| 8.6 |  | 37 |
| 8.7 |  | 38 |
| 8.8 |  | 39 |
| 8.9 |  | 40 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **TABLE NO** | **NAME OF THE TABLE** | **PAGE NUMBER** |
| 7.1 |  | 32 |
| 7.2 |  | 32 |
| 8.1 |  | 38 |
| 8.2 |  | 39 |
| 8.3 |  | 40 |

**LIST OF ABBREVIATION**

|  |  |
| --- | --- |
| **ABBREVIATION** | **EXPANSION** |
| KNN | K- Nearest Neighbor Classifier |
| CNN | Convolutional Neural Network |
| API | Application Programming Interface |
| PCA | Principal Component Analysis |
| XGBoost | Gradient Boosting |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

1. **INTRODUCTION**

Patient medical records are captious in the healthcare field for predicting variety of disease and determining the huge impact. Organizing and keeping track of all history of the patient becoming tedious and priority for treatments which requires more significance.

To administer the above problem, this model focusses on providing better solution by using Ensemble Machine Learning algorithms, KNN, Logistic Regression, Random Forest, XGBoost etc., to increase the performance and optimization of classification and prediction of patient data and produce the results based on the prediction of various diseases.

**DATA PREPROCESSING**

Data Preprocessing is one of the data mining Technique which is used to transform the raw data in a functional and efficient format. It makes the more data analysis or visualization more easier. All missing values, Redundant values, unreliable data can be preprocessed and filled by means of mean or median. The final Outcome of data preprocessing is turned as Training Dataset.

**MACHINE LEARNING**

Machine learning (ML) is the process of using mathematical models of data to help a computer learn without direct instruction. It’s considered a subset of artificial intelligence (AI). Machine learning uses algorithms to identify patterns within data, and those patterns are then used to create a data model that can make predictions. With increased data and experience, the results of machine learning are more accurate much like how humans improve with more practice.

DEEP LEARNING

Deep learning is a subset of machine learning which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to “learn” from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

* 1. **PROBLEM STATEMENT**

The healthcare industry has become big business. The healthcare industry produces large amounts of healthcare data every day that can be used to extract information to predict diseases that a patient may experience in the future while using treatment history and healthcare data. This information hidden in health data is later used to make affective decisions for the patient's health. In addition, this area needs to be improved through the use of meaningful data in healthcare. The rapid adoption of electronic health records has created a wealth of new patient data that represents a gold mine for a better understanding of human health. The main challenge is to extract the information from this data as the amount is very large so some data mining and machine learning techniques can be used.

During the manual procedure, the doctor's first step is to determine whether the disease is benign or malignant and to evaluate the blood values and perform a biopsy of the specific part. Patients may or may not have collective ailments that are not identified when the procedure is performed concurrently. To increase efficiency in predicting the types of diseases people suffer from and finding the multitude of diseases that are re-emerging in the environment, Custom Ensemble Learning is used with Pipeline, KNearest Neighbor (KNN), Logistic Regression, Convolutional Neural Networks (CNN) and Principal Component uses analysis. In addition, the expected outcome and scope of this project is that if the disease can be predicted, patients can be treated early, which can reduce life risk and save patients' lives, and the cost of disease treatment can be partially reduced through early detection.

1. **LITERATURE REVIEW**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Author** | **Disease** | **Research Objective** | **Algorithm** | **Accuracy** |
| **1.** | (Naqi et al.,  2020) | Lung  cancer | The system's problem includes false positive results, this work provides an automated detection system and classification to promote radiologists' diagnosis. |  | 96.9% |
| 2 | (Shivangi et al.,2019) | Parkinson’s Disease | This paper proposes two neural network-based models which aims at early diagnosing of the disease, thus proposed models outperforming the existing methodology in terms of accuracy. |  | 89.15% & 88.1% |
| 3 | (Javeed et al.,  2019) | Heart  disease | Developed an intelligent system that would show good performance on both training and testing data diagnosis of heart failure. |  | 93.33% |
| 4 | (Durai et al.,  2019) | Liver  disease | Predicted the accurate and definitive result,  compared algorithm techniques with a  higher accuracy rate for detecting liver disease. Along with the introduction of widely analyzed classification model consisting of five distinct process for the same. |  | 95.04% |
| 5 | (Kumar et al.,  2020) | Blood  Cancer | Author proposed DL techniques, namely CNNs, the proposed model eradicates the manual method's likelihood of errors. The model, trained on cells’ images, preprocesses the images first and extracts the best characteristics. | CNN | 97.2% |
| 6 | (R Sharmila et al., 2018) | Heart Disease | Author proposed a conceptual method to enhance the method of heart disease using data mining techniques. | SVM | 85% |
| 7 | (Liu et al., 2019) | Brain Stroke | The expanding of hematoma is in anticipation that spontaneously ICH derives from accessible compared by the usage of SVM. | SVM | 83.3% |
| 8 | (Qawqzeh et al., 2020) | Diabetes | Author proposed a logistic regression model based on  Photoptheysmogram analysis for diabetes classification. |  | 92% |
| 9 | (Ahmed et al., 2019) | Alzheimer disease | The objective is to increase the degree of accuracy comparable to the problem of overfitting and examine validated brain technologies. |  | 90.05% |
| 10 |  |  |  |  |  |

1. **SYSTEM ANALYSIS**

**3.1 EXISTING SYSTEM**

Artificial Intelligence is a buzz word in this century that is the basis of a lot of computing projects to be applied to very different parts of our life. Early-stage detection of diseases is very essential and necessary because of the unhealthy food habits and lifestyle. In the study for predicting disease and to classify the disease based on accuracy which determines the person is healthy or not. Machine Learning is used to identify, detect and predict the accuracy of various diseases. KNN, Linear Regression, Naïve Bayes, Decision Tree, Neural Networks was implemented to carry out the detection and prediction processes very precisely. Model which was built using KNN attained 92% accuracy in prediction of Cancer Disease. Neural Network models built which aims at early diagnosing of the disease.

**3.2 PROPOSED SYSTEM**

The cardinal objective of the project is to develop new model of disease diagnosis using Custom ensemble, K-Nearest Neighbor(KNN),Naïve Bayes, Decision, Tree, Convolutional Neural Networks(CNN), Random Forest Classifier infused together for high performance-oriented outcomes in terms of disease diagnosis with quick and accurate results to the maximum possible.

The model is well trained and tested on the dataset so as to attain the best accuracy outcomes. The model can be designed to detect the Disease which finds out whether it’s benign or malignant through appropriate check of entire body’s parameters. Machine learning algorithms models are designed for detect the disease stages and predicts the stages in a stepwise manner. The inclusion of Custom ensemble learning techniques thus enabling the improvement and boost up of the overall performance of the models developed.

1. **SYSTEM SPECIFICATIONS**

**4.1 SOFTWARE REQUIREMENTS**

* Python
* Libraries such as numpy, pandas, sklearrn, seaborn, Flask
* Jupyter notebook or any python code editor
* Visual Studio Code

**4.2 HARDWARE REQUIREMENTS**

* Intel i5 7th gen
* 500gb Disk Space
* Nvidia GeForce 1050Ti

1. **SOFTWARE DESCRIPTION**

**PYTHON**

Python is a Open-Source programming language which is easy to read and very powerful. The main applications of python such as Machine learning, Web development. In the project, Python is widely used to do more calculations, to create machine learning models, imported many scientific and mathematical libraries to predict accuracy by the means of algorithms.

**FLASK**

Flask is a web framework, it’s a Python module that lets you develop web applications easily. It’s has a small and easy-to-extend core: Flask is based on the Werkzeg WSGI toolkit. Flask is used in the project to build the web based application embedding with machine learning to predict the accuracy at the ease.

**JUPYTER NOTEBOOK**

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people. Jupyter supports many languages that is Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python. In the project, Jupyter notebook is used to run Python scripts and machine learning models in simplest way.

1. **PROJECT DESCRPTION**

**6.1 Problem Definition**

**6.2 Introduction to the Proposed System**

**6.3 MODULE DESCRIPTION**

**6.3.1 Importing Modules and Dataset**

**Importing Library**

On importing several libraries for the project:

1. Numpy: NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

2. Pandas: To work with csv files and dataframespandas is a software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.

3. Matplotlib: To create charts using pyplot, define parameters using rcParams and color them with cm.rainbow

4. Warnings: To ignore all warnings which might be showing up in the notebook due to past/future depreciation of a feature

5. train\_test\_split: To split the dataset into training and testing data

6. StandardScaler: To scale all the features, so that the Machine Learning model better adapts to the dataset.

7. Seaborn: Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions

8. Sklearn: Scikit-learn (Sklearn) is the most useful robust library for machine learning in Python for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

**Importing Dataset**

Before preprocessing the raw dataset, Importing dataset is first and foremost step. Datasets are fetched from Kaggle. Raw datasets contains missing values, Nan values etc., Thorough Analysis is required to process the dataset. To fetch the attributes of the datasets, many methods are there like head(), tail(), describe(), info()

Text

Description automatically generated

Fig 6.1 Importing datasets and fetching using head()

Text

Description automatically generated Datasets consists of various parameters (independent variables and one target (dependent variable, Outcome, Independent variables includes the age, BMI, density, insulin level, sex, number of pregnancies and so on.

Fig 6.2 Description Statistics of the dataset done using describe()

Text

Description automatically generated**6.3.2 Exploratory Data Analysis**

Fig 6.3 Exploratory data analysis of heart disease dataset

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

**6.3.3 Data Pre-Processing**

Data Preprocessing is one of the data mining Technique which is used to transform the raw data in a functional and efficient format. It makes the more data analysis or visualization more easier. All missing values, Redundant values, unreliable data can be preprocessed and filled by means of mean or median. The final Outcome of data preprocessing is turned as Training Dataset.

Text

Description automatically generatedTo fill the missing values and unreliable data in the dataset, Mean or median or other preprocessing techniques like exploratory data analysis and Statistical data analysis

Fig 6.3 Filling missing values using preprocessing methods

df is called as dataframes which is the variable assigned to fetch dataset

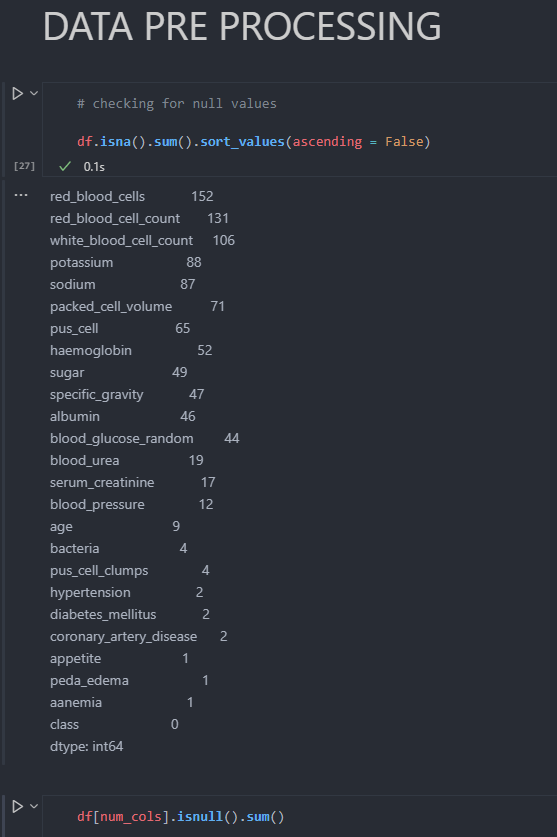
****

Fig 6.4 Finding the total sum of the missing values

Isnull() is a method used to find whether the dataset is having null columns and sum() is used to count the iterations of missing or null columns.

Isna() is a method used to find Not available values in the columns of the dataset.

Preprocessed data is now ready to get trained and tested by the means of various algorithms.

**6.3.4 Detection using KNN, Regression, XGBoost, Decision Tree**

**KNN algorithm**

KNN algorithm predicts the class of a new instance based on the most votes by its closest neighbors. It uses Euclidean distance to calculate the distance of an attribute from its neighbors. This classifier looks for the classes of K nearest neighbors of a given data point and based on the majority class, it assigns a class to this data point. However, the number of neighbors can be varied. It varied from 1 to 20 neighbors and calculated the test score in each case. And mostly this is the simplest and efficient way for diagnosing the heart disease analysis. This system has predicted 87 per cent accurately more than the expert’s Statistical analysis



**Logistic Regression**

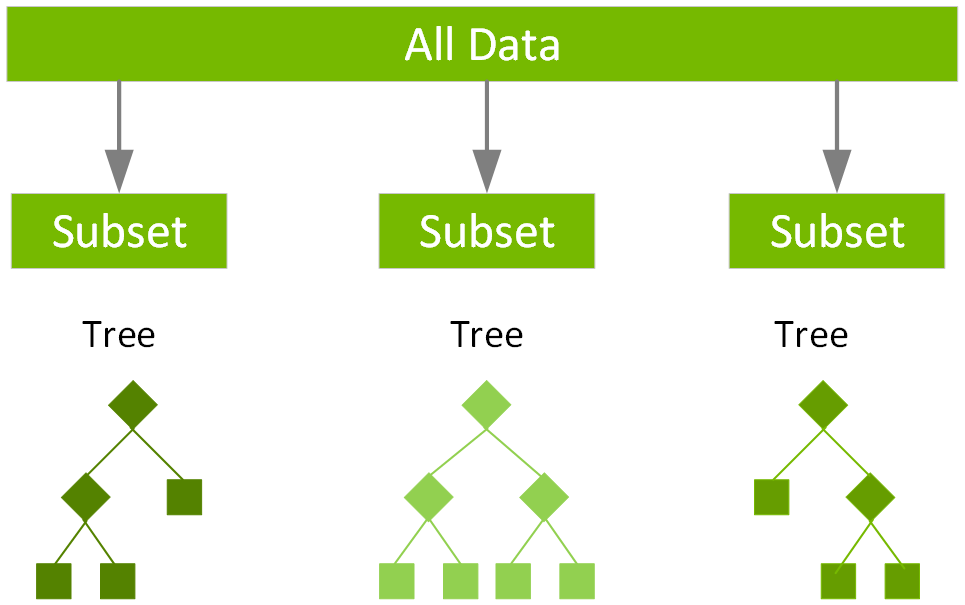
Logistic regression is a classification algorithm that is applied in situations when the output variable is *categorical*. The goal of Logistic Regression is to discover a relationship between features and the probability of a specific outcome. Logistic Regression’s gradient descent algorithm will look identical to Linear Regression’s gradient descent algorithm. For the case of gradient descent, the search direction is the negative partial derivative of the logistic regression cost function with respect to the parameter θ. In its most basic form, gradient descent will iterate along the negative gradient direction of θ (known as a *minimizing sequence*) until reaching convergence.

**log[p(X) / (1-p(X))]  =  β0 + β1X1 + β2X2 + … + βpXp**

where: **Xj**: The jth predictor variable **βj**: The coefficient estimate for the jth predictor variable

then use some probability threshold to classify the observation as either 1 or 0.

**XGBoost**

A Gradient Boosting Decision Trees (GBDT) is a decision tree ensemble learning algorithm similar to random forest, for classification and regression. Ensemble learning algorithms combine multiple machine learning algorithms to obtain a better model. The term “gradient boosting” comes from the idea of “boosting” or improving a single weak model by combining it Text

Description automatically generatedwith a number of other weak models in order to generate a collectively strong model.

Fig 6.6 models comparison with different algorithms

**6.3.5 DETECTION USING CUSTOM ENSEMBLE LEARNING**

**6.3.6 WEBSITE DEPPLOYMENT**

**Website Deployment using Flask**

Flask is a web framework, it’s a Python module that lets you develop web applications easily. It’s has a small and easy-to-extend core: Flask is based on the Werkzeg WSGI toolkit. Flask is used in the project to build the web based application embedding with machine learning to predict the accuracy at the ease.

After designing of model to predict the accuracy of the diseases, Flask is used with python and HTML to build disease detection website which detects, classifies, predict the accuracy and also tells whether the person is infected with particular disease. The model is built using both machine learning and deep learning. In the website, pictures of the image samples taken during the tests must be uploaded to get the results.

**7. SYSTEM IMPLEMENTATIONS**

**7.1 DISEASE DETECTION USING MACHINE LEARNING**

**7.1.1 Detection using Custom Ensemble Machine Learning**

**7.1.2 Detection using Random Forest**

Detection using Random Forest Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forest uses a technique called bagging to build full decision trees in parallel from random bootstrap samples of the data set. The final prediction is an average of all of the decision tree predictions.

**7.2 DISEASE DETECTION USING DEEP LEARNING**

**7.2.1 Detection using Convolutional Neural Network(CNN)**

**8. RESULTS AND DISCUSSIONS**

8.1 Experimental Results

8.2 Comparison of Results

**9.CONCLUSION AND ENHANCEMENTS**

**10. REFERENCES**

[1] Syed Muhammad Naqi,Muhammad Sharif,Mussarat Yasmin, “Multistage segmentation model and SVM-ensemble for precise lung nodule detection”,International Journal of Computer Assisted Radiology and Surgery", Vol 13, February 2018.

[2] Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S. C., Senkerik, R., & Hemanth, J. D. “Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using Convolutional Neural Networks. IEEE Access, 8(Mm), 142521–142531,2020.

[3] Vasan Durai,Suyan Ramesh(1), “Liver disease prediction suing Machine learning”, International Journal of Advance Research, Ideas and Innovations in Technology[IJARIIT],Vol 5,Issue 2, 2019.

[4] Liaqat Ali,Atiqur Rahman,Aurangazeb Khan,Mingyi Zhou,Ashir Javeed and Javed Ali Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on χ 2 Statistical Model and Optimally Configured Deep Neural Network”, IEEE Access, Vol 7, April 2019.

[5] Shivangi, Anubhav Johri,Ashish Tripathi, “Parkinson Disease Detection Using Deep Neural Networks”, IEEE, Twelfth International Conference on Contemporary Computing(IC3),2019.

[6] M. Islam, H. Iqbal, R. Haque, and K. Hasan, “Prediction of breast cancer using support vector machine and K-nearest neighbors,” in Proceedings of the IEEE Region 10 Humanitarian Technology Conference (R10-HTC), vol. 23, pp. 1–5, Dhaka, Bangladesh, December 2017.

[7] Al-Zebari, A., & Sengur, A. (2019). Performance Comparison of Machine Learning Techniques on Diabetes Disease Detection.1st International Informatics and Software Engineering Conference: Innovative Technologies for Digital Transformation, IISEC 2019.

[8] D. Zhang, L. Zou, X. Zhou, and F. He, “Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer,” IEEE Access , vol. 6, pp. 28936–28944, 2018.

[9] E. Alickovi ˇ c and A. Subasi, “Breast cancer diagnosis using GA ´ feature selection and rotation forest,” Neural Computing and Applications, Vol. 28, no. 4, pp. 753–763, 2017.

[10] A. U. Haq, J. P. Li, M. H. Memon et al., “Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson’s disease using voice recordings,” IEEE Access, vol. 7, pp. 37718–37734, 2019.

**APPENDIX**

**CODE**