ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/370399643

Developing a Machine Learning-Based Multiple Disease Prediction System: A Comprehensive Analysis of Risk Factors and Disease Interactions

Research · April 2023

CITATIONS
0

READS
10

4 authors, including:

Emad Naushad
ITS Engineering College
1 PUBLICATION **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Emad Naushad on 29 April 2023.

The user has requested enhancement of the downloaded file.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

Developing a Machine Learning-Based Multiple Disease Prediction System: A Comprehensive Analysis of Risk Factors and Disease Interactions

Emad Naushad[1], Bhavishya Raj[2], Arpit Nirvan[3], Vrinda Sachdeva[4]
Students, B.Tech Computer Science & Engineering[1,2,3]
Associate Professor of Department of Computer Science & Engineering[4]
I.T.S Engineering College, Greater Noida, India

Abstract: *Using predictive modelling, the "Multiple Disease Prediction System" foretells the user's sickness depending on the symptoms are supplied as input to the system. The system evaluates the user's symptoms as input and outputs the likelihood that the disease will occur. The Random Forest Classifier is used for prediction, and Deep Learning Models for Diabetes, Heart Disease and Parkinson's Disease. This method is more accurate and a construction of a web application for prediction system is done.*

Keywords: Parkinson's Disease, Predictive Modelling, Diabetes, Heart Disease

I. INTRODUCTION

Anyone who is currently ill must see a doctor, which is both time consuming and costly. Because the sickness cannot be identified, it might also be challenging for the user if it is out of reach of doctors and hospitals. So, if the following treatment can be performed using automated software that saves time and money, it may be better for the patient and make the process move more smoothly. Other Multiple Disease Prediction Systems examine the patient's risk level using data mining approaches. ailment Predictor is a web-based application that predicts a user's ailment based on their symptoms. For the Disease Prediction system, data sets from several health-related websites were acquired.Using condition Predictor, the customer will be able to predict the possibility of a condition based on the symptoms provided. People are constantly eager to learn new things, especially as the use of the internet expands by the day. When a problem arises, individuals frequently want to search it up on the internet. Hospitals and physicians have less internet connectivity than the general public. People who are plagued with a sickness do not have many options. As a result, people may benefit from this system. Chronic illness is a disease that lasts for an extended period or takes a long time to heal, and many chronic diseases cannot be cured but must be managed daily India, like all other countries, is undergoing substantial social and economic changes, which is leading to an increase in the prevalence of cardiovascular disease. Many established, developing, and developing countries, including India, are coping with a wide spectrum of chronic diseases, particularly cardiovascular disease, which has major implications for global health, security, and the economy. The world's growing urbanisation and economic progress have resulted in a diverse spectrum of lifestyles. Chronic diseases are now an issue in all countries, affecting one-third of the population in each. Chronic disease treatment is more expensive, and it is challenging on the sick. A vast number of chronic disease datasets are collected and processed in the medical area, and data mining aids in disease early detection. The most expensive diseases to diagnose include heart disease, diabetes, and Parkinson's disease.

Offering the finest quality services to all patients is a huge difficulty in the medical or healthcare industry, and only those who can afford it may profit from it. There is a tremendous amount of healthcare data available that is not being mined in a more efficient and dependable manner to unearth hidden knowledge for effective decision-making. To diagnose chronic diseases early, the suggested system leverages data mining approaches. Machine learning is the process of teaching computers to improve their output based on past data or examples. Machine learning is the study of computer systems that learn from data and experience. The machine learning algorithm has two stages: training and testing. Prediction of a disease based on the patient's symptoms and medical history for decades, machine learning has

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

been a stumbling barrier. In the medical sector, Machine Learning technology provides a strong venue for rapidly resolving healthcare challenges.

II. RESEARCH OBJECTIVE

There is a need to explore and build a system that would allow end users to forecast chronic diseases without needing to consult a physician or doctor. Identifying various diseases by studying patients' symptoms and employing various Machine Learning Models approaches. There is no standard process for dealing with text and structured data. The proposed approach would consider both organised and unstructured data. Machine Learning can enhance forecast accuracy.

III. LITERATURE REVIEW

Diabetes is a chronic metabolic disorder that affects millions of people in India. According to the International Diabetes Federation, India had 77 million adults (aged 20-79 years) living with diabetes in 2019, and this number is projected to rise to 134 million by 2045 [1]. Literature on diabetes in India indicates that there are various risk factors that can contribute to the development of this disease, including obesity, physical inactivity, and genetics [2]. In addition to these risk factors, the prevalence of diabetes is also associated with socioeconomic status, age, and ethnicity in India [3]. The management of diabetes in India involves lifestyle changes such as diet and exercise, along with medications such as metformin and insulin [4]. In recent years, research in India has also focused on the use of traditional Indian medicines, such as Ayurveda, to manage diabetes [5]. These therapies hold promise for the treatment of diabetes in India, where traditional medicine plays an important role in healthcare. Heart disease is another chronic disease that affects millions of people in India. According to the Indian Heart Association, cardiovascular disease CVD is the leading cause of death in India, responsible for 28% of all deaths [6]. Literature on heart disease in India indicates that there are several risk factors that can contribute to its development, including smoking, high blood pressure, and high cholesterol [7]. In addition to these risk factors, the prevalence of heart disease is also associated with age, sex, and genetics in India [8]. The management of heart disease in India involves lifestyle changes such as diet and exercise, along with medications such as statins and antiplatelet drugs [9]. In recent years, research in India has also focused on the use of traditional Indian medicines, such as Ayurveda, to manage heart disease [10]. These therapies hold promise for the treatment of heart disease in India, where traditional medicine is an integral part of the healthcare system. Parkinson's disease is a progressive neurological disorder that affects millions of people in India. According to a study published in the Journal of Parkinson's Disease, the prevalence of Parkinson's disease in India is 39.4 per 100,000 population [11]. Literature on Parkinson's disease in India indicates that there are several risk factors that can contribute to its development, including genetics, environmental factors, and aging [12]. In addition to these risk factors, the prevalence of Parkinson's disease is also associated with sex, with men being more likely to develop the disease than women in India [13]. The management of Parkinson's disease in India involves medications such as levodopa and dopamine agonists, along with physical therapy and deep brain stimulation [14]. In recent years, research in India has also focused on the use of traditional Indian medicines, such as Ayurveda, to manage Parkinson's disease [15]. These therapies hold promise for the treatment of Parkinson's disease in India, where traditional medicine is an important part of the healthcare system.

This research paper was written by Emad Naushad, Bhavishya Raj, Arpit Nirvan andVrinda Sachdeva to provide a survey of existing techniques of information discovery in databases using data mining techniques that are used in today's medical research, specifically in Multiple Disease Prediction. Several experiments have been carried out to compare the performance of predictive modelling techniques on the same dataset, and the results show that Decision Tree outperforms, with Bayesian classification having comparable accuracy to Decision Tree in some cases, but other predictive approaches such as SVM, Logistic Regression, and Classification based on Clustering underperform.

A study was conducted to predict heart diseases using the Decision Tree Algorithm, in which the consumer provides data that is compared to a qualified set of values. As a result of this study, patients were able to provide basic information that was compared to data, and heart disease was expected.Also, analysis of the various types of heart-related problems using medical data mining techniques such as association rule mining, grouping, and clustering I. The aim of a decision tree is to show any possible outcome of a decision. To achieve the best result, various rules are

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

devised. The criteria used in this study were age, sex, smoking, being overweight, drinking alcohol, blood sugar, heart rate, and blood pressure.

IV. PROPOSED SYSTEM

Architecture Diagram

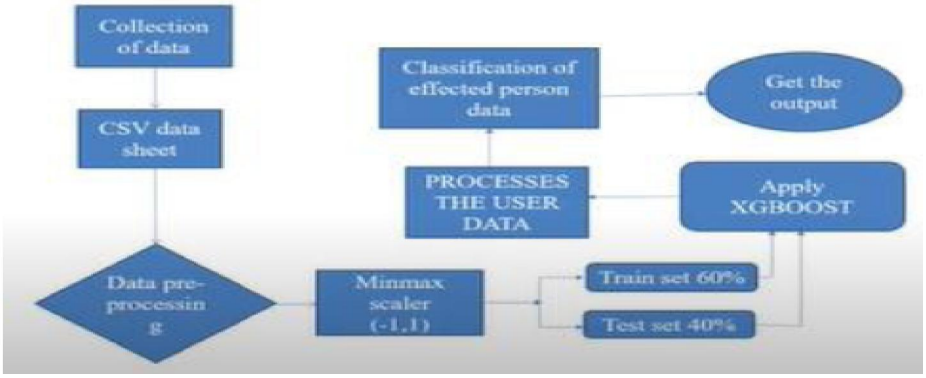


Fig.1. Architecture Diagram

Work Flow Diagram

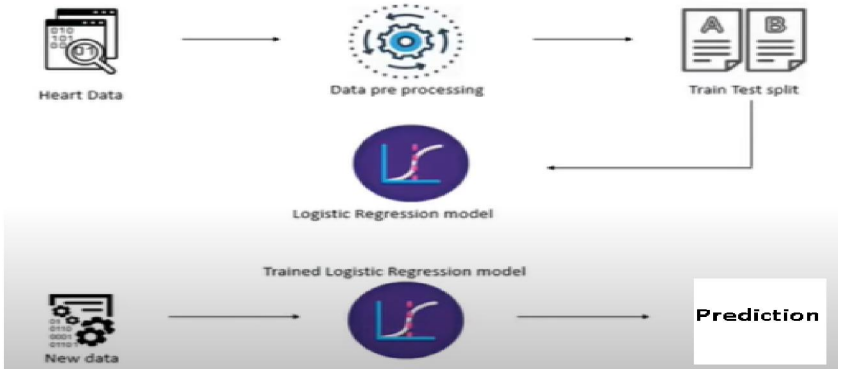
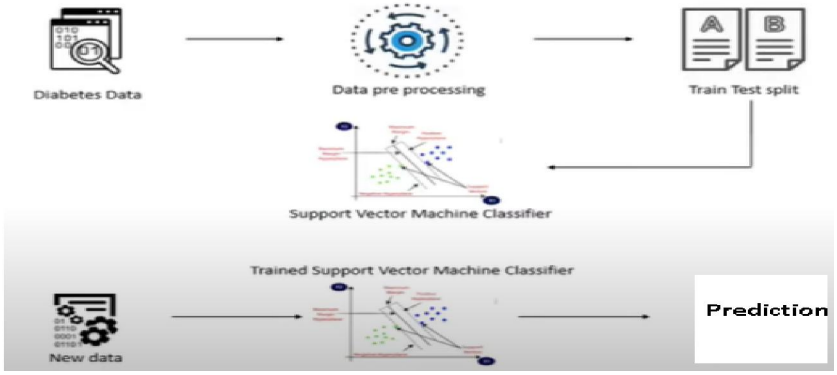


Fig.2. Work Flow Diagram

4.1 Data Collection



Data collection has been done from the internet to identify the disease here the real symptoms of the disease are collected i.e., no dummy values are entered. The symptoms of the disease are collected from different health related websites.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

4.2 Data Pre-Processing

Before feeding the data into the Prediction model, following data cleaning and pre-processing steps are performed
Checking null values and filling using forward fill method

4.3 Building Model

- Converting data into different cases
- Standardizing the data using mean and standard deviation
- Splitting the dataset into training and testing sets

Many methods are used to perform data mining. Machine learning is one of the approaches. Random forest Machine learning strategies include grouping, clustering, summarization, and many others. Since classification techniques are used in this project, classification is one of the data mining processes in this phase of categorical data classification. And this step is divided into two phases: training and testing. In the training phase, predetermined data and associated class labels are used for classification. The training stage is often referred to as supervised learning. The preparation and testing phases of the classification process are depicted in the diagram. In the training process, training tuples are used, and in the test data phase, test data tuples are used, and the classification rule's accuracy is calculated. Assume that the classification rule's accuracy on testing data is sufficient for the rule to be used for classification of unmined data.

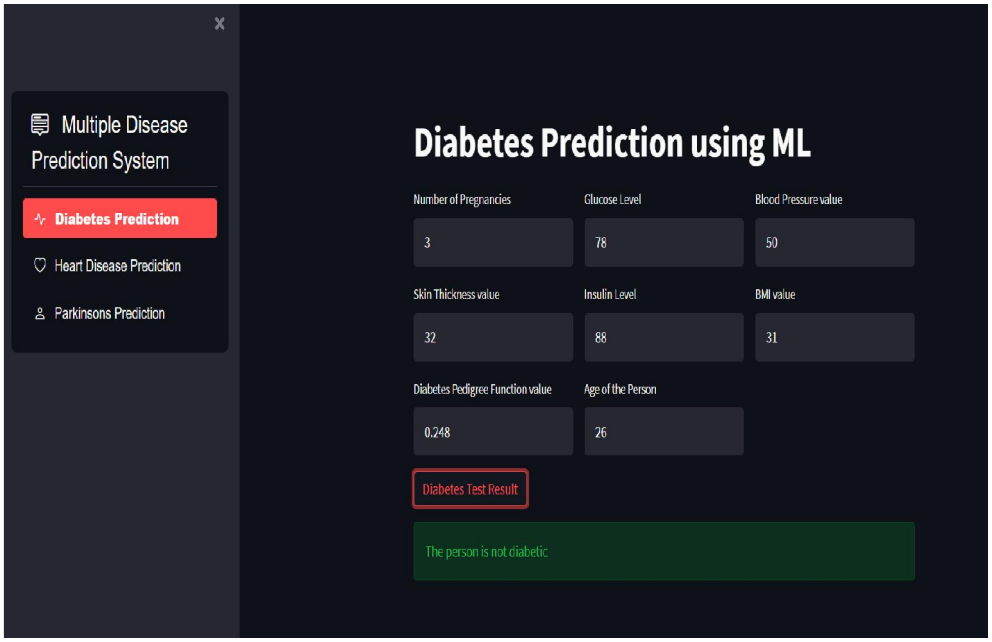
4.4 Prediction

Prediction using Random Forest: -
Prediction done by Random Forest Model using Streamlit framework model trained by training chronic disease dataset.

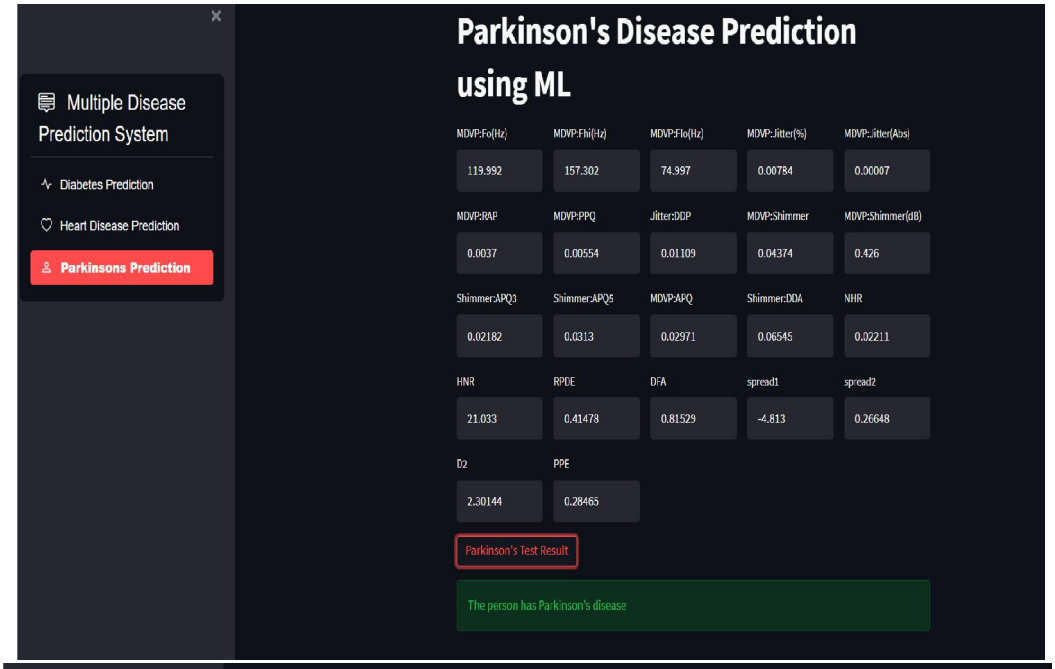
4.5 Algorithm

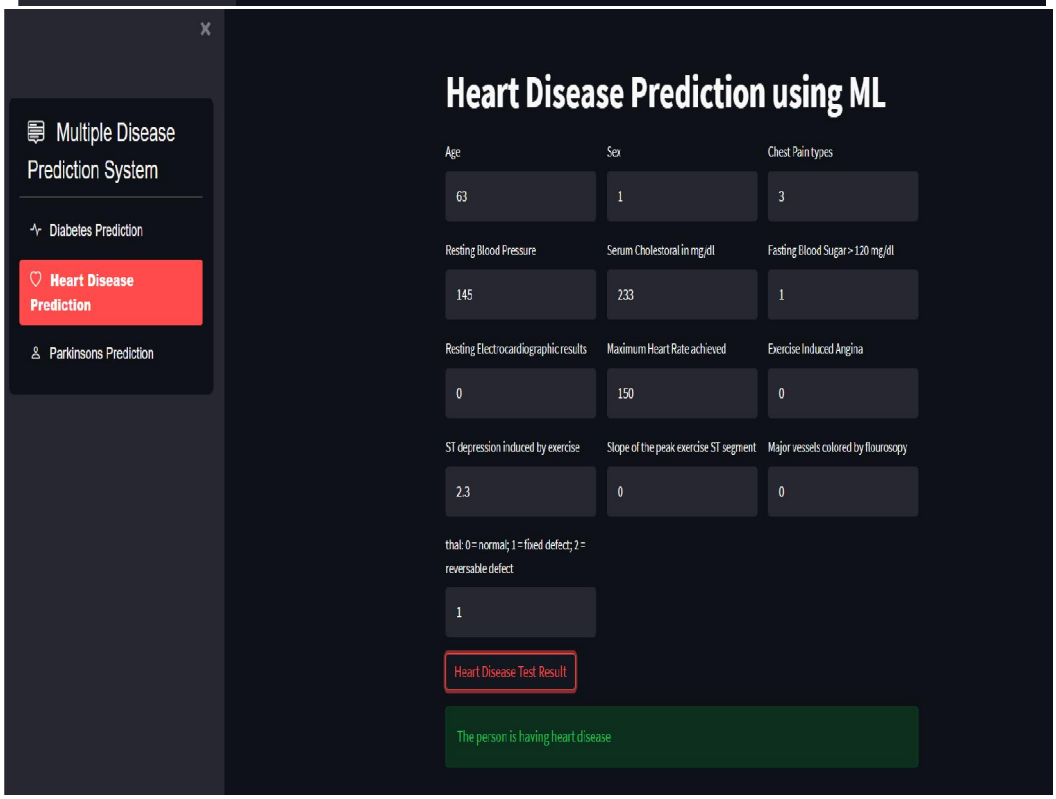
Logistic regression predicts the output of a categorical dependent variable.
Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
SVM (Support Vector Machine) algorithm creates the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.This best decision boundary is called a hyperplane.

V. OUTCOMES



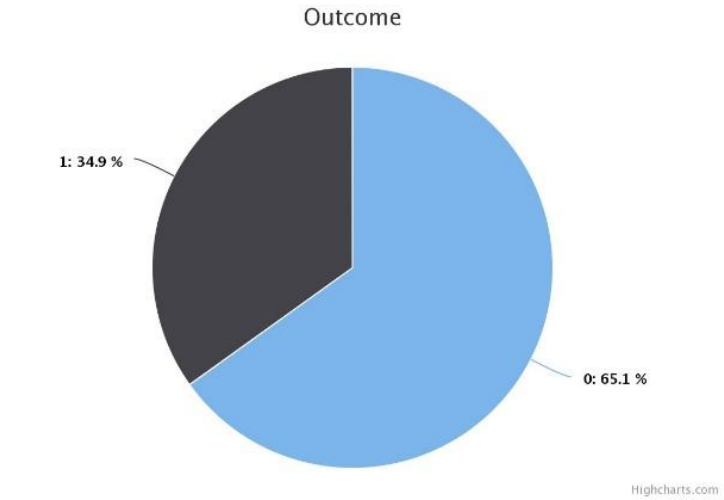
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

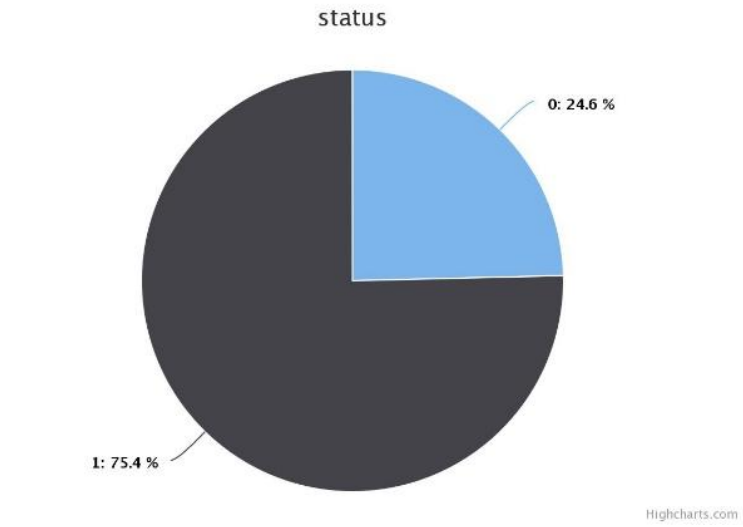


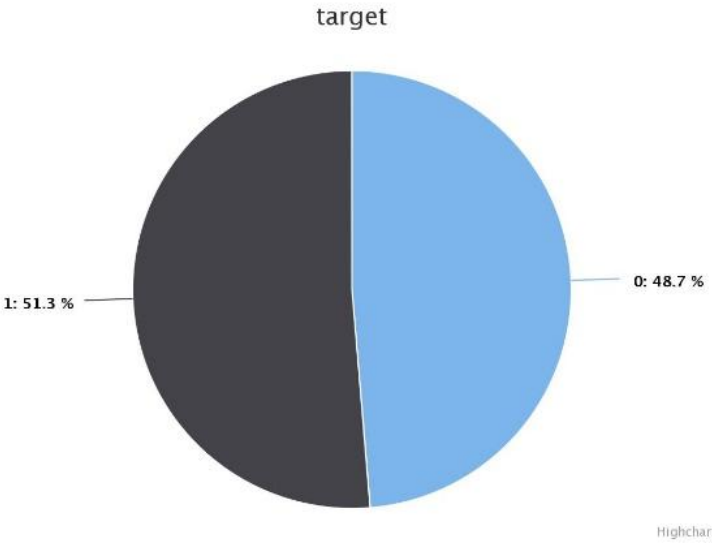


IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

VI. CHARTS







IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

VII. DATA

	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0
23	8	99	84	0	0	35.4	0.388	50	0
24	7	196	90	0	0	39.8	0.451	41	1
25	9	119	80	35	0	29	0.263	29	1
26	11	143	94	33	146	36.6	0.254	51	1
27	10	125	70	26	115	31.1	0.205	41	1
28	7	147	76	0	0	39.4	0.257	43	1
29	1	97	66	15	140	23.2	0.487	22	0

Diabetes dataset

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
2	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01105	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.81303	0.266482	2.301442	0.284654
3	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.07519	0.33559	2.486855	0.368674
4	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	1	0.429895	0.825288	-4.44318	0.311173	2.342259	0.332634
5	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.03772	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.1175	0.334147	2.405554	0.368975
6	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	1	0.417356	0.823484	-3.74779	0.234513	2.33218	0.410335
7	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.415564	0.825069	-4.24287	0.299111	2.18756	0.357775
8	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01603	0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.886	1	0.59604	0.764112	-5.63432	0.257682	1.854785	0.211756
9	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.63742	0.763262	-6.1676	0.183721	2.064693	0.163755
10	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02099	0.191	0.01073	0.01277	0.01717	0.03218	0.0107	21.812	1	0.615551	0.773587	-5.49868	0.327769	2.322511	0.231571
11	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	1	0.547037	0.798463	-5.01188	0.325996	2.432792	0.271362
12	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	1	0.611137	0.776156	-5.24977	0.351002	2.407313	0.24974
13	91.904	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.02214	0.04272	0.01141	21.414	1	0.58339	0.79252	-4.96023	0.363566	2.642476	0.275931
14	136.926	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968	0.00581	25.703	1	0.4606	0.646846	-6.54715	0.152813	2.041277	0.138512
15	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01797	0.02184	0.01041	24.889	1	0.430166	0.665833	-5.66022	0.254989	2.519422	0.199889
16	152.845	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364	0.01823	0.158	0.01064	0.00972	0.01246	0.03191	0.00609	24.922	1	0.474791	0.654027	-6.1051	0.203653	2.125618	0.1701
17	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.01359	0.02316	0.00839	25.175	1	0.565924	0.658245	-5.34012	0.210185	2.205546	0.234589
18	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.02074	0.02908	0.01859	22.333	1	0.56738	0.644692	-5.44004	0.229764	2.264501	0.213164
19	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322	0.02919	20.376	1	0.631099	0.605417	-2.93107	0.434326	3.007463	0.430788
20	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413	0.0316	17.28	1	0.665318	0.719467	-3.94908	0.25787	3.10901	0.377429
21	156.405	189.358	142.822	0.00768	0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164	0.03365	17.153	1	0.649554	0.68608	-4.55447	0.340176	2.856676	0.322111
22	153.848	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05	0.03871	17.536	1	0.660125	0.704087	-4.09544	0.262564	2.73971	0.363391
23	153.88	172.86	78.128	0.0048	0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.04525	0.06062	0.01849	19.493	1	0.629017	0.698951	-5.18696	0.237622	2.557536	0.259765
24	167.93	193.221	79.068	0.00442	0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.04246	0.06685	0.0128	22.468	1	0.61906	0.679834	-4.33096	0.262384	2.916777	0.285695
25	173.917	192.735	86.18	0.00476	0.00003	0.00221	0.00258	0.00663	0.04192	0.364	0.02187	0.0247	0.03772	0.06562	0.0184	20.422	1	0.537264	0.686894	-5.24878	0.210279	2.547508	0.259556
26	163.656	200.841	76.779	0.00742	0.00005	0.0038	0.0039	0.0114	0.01659	0.164	0.00738	0.00948	0.01497	0.02214	0.01778	23.831	1	0.397937	0.732479	-5.55745	0.22089	2.692176	0.215961
27	104.4	206.002	77.968	0.00633	0.00006	0.00316	0.00375	0.00948	0.03767	0.381	0.01732	0.02245	0.0378	0.05197	0.02887	22.066	1	0.522746	0.737948	-5.57184	0.236853	2.846369	0.213514
28	171.041	208.313	75.501	0.00455	0.00003	0.0025	0.00234	0.0075	0.01965	0.186	0.00889	0.01169	0.01872	0.02666	0.01095	25.908	1	0.418622	0.720916	-6.18359	0.226278	2.589702	0.147403
29	146.845	208.701	81.737	0.00496	0.00003	0.0025	0.00275	0.00745	0.01913	0.198	0.00883	0.01144	0.01826	0.0265	0.01328	25.119	1	0.358773	0.726652	-6.27169	0.156102	2.314209	0.162999
30	155.358	227.383	80.055	0.0031	0.00002	0.00159	0.00176	0.00476	0.01713	0.161	0.00769	0.01012	0.01661	0.02307	0.00677	25.97	1	0.470478	0.676258	-7.12059	0.279789	2.241742	0.108514
31	162.568	198.346	77.63	0.00502	0.00003	0.0028	0.00253	0.00841	0.01791	0.168	0.00793	0.01057	0.01799	0.0238	0.0117	25.678	1	0.427785	0.723797	-6.63573	0.269866	1.957961	0.135242
32	197.076	206.896	192.055	0.00289	0.00001	0.00166	0.00168	0.00498	0.01099	0.097	0.00563	0.0068	0.00802	0.01689	0.00339	26.775	0	0.422229	0.741367	-7.3483	0.177551	1.743867	0.085569
33	199.228	209.512	192.091	0.00241	0.00001	0.00134	0.00138	0.00402	0.01015	0.089	0.00504	0.00641	0.00762	0.01513	0.00167	30.94	0	0.432439	0.742055	-7.68259	0.173319	2.103106	0.068501
34	198.383	215.203	193.104	0.00212	0.00001	0.00113	0.00135	0.00339	0.01263	0.111	0.0064	0.00825	0.00951	0.01919	0.00119	30.775	0	0.465946	0.738703	-7.06793	0.175181	1.512275	0.09632
35	202.266	211.604	197.079	0.0018	0.00009	0.00093	0.00107	0.00276	0.00954	0.085	0.00469	0.00606	0.00719	0.01407	0.00072	32.684	0	0.368535	0.742133	-7.69573	0.17854	1.544609	0.055141
36	203.184	211.526	196.16	0.00178	0.00009	0.00094	0.00106	0.00283	0.00959	0.085	0.00468	0.0061	0.00726	0.01403	0.00065	33.047	0	0.340068	0.741899	-7.96498	0.163519	1.423287	0.044539
37	201.464	210.565	195.708	0.00198	0.00001	0.00105	0.00115	0.00314	0.01194	0.107	0.00586	0.0076	0.00957	0.01758	0.00135	31.732	0	0.344252	0.742737	-7.77769	0.170183	2.447064	0.05761

Parkinson dataset

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-9588

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1	
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1	
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1	
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1	
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1	
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1	
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1	
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1	
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1	
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1	
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1	
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1	
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1	
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1	
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1	
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1	
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1	
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1	
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1	
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1	
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1	
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1	
24	42	1	0	140	226	0	1	178	0	0	2	0	2	1	
25	61	1	2	150	243	1	1	137	1	1	1	0	2	1	
26	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1	
27	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1	
28	59	1	2	150	212	1	1	157	0	1.6	2	0	2	1	
29	51	1	2	110	175	0	1	123	0	0.6	2	0	2	1	

Heart disease dataset

VIII. CONCLUSION

The proposed work brings diabetes, heart disease, and Parkinson disease under a single platform by deploying the trained models using the Streamlit framework which is a lightweight framework. One classification and one regression algorithms are used for training the models, in which the SVM gave good accuracy values for the disease prediction of diabetes and Parkinson Logistic regression for the disease prediction of heart disease. Its highest accuracy is calculated by picking the highest value obtained from 1 to 21 neighbours. In the future, we can expand this work by adding more diseases that are trained by machine learning models and can include the disease that involves deep learning models.

REFERENCES

[1]. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. Int J Epidemiol. 2011;40:804–818.

[2]. Kalia, L.V.; Lang, A.E. Parkinson's Disease. Lancet 2015, 386, 896–912. 4. D. Heisters, "Parkinson's: symptoms treatments and research", vol. 20, no. 9, pp. 548-554, 2011.

[3]. Zhilbert Tafa, Nerxhivane Pervetica, Bertran Karahoda, "An Intelligent Systemfor Diabetes Prediction", 4thMediterranean Conference on EmbeddedComputing MECO – 2015 Budva, Montenegro

[4]. Mahlknecht, P.; Krismer, F.; Poewe, W.; Seppi, K. Meta-Analysis of Dorsolateral Nigral Hyperintensity on Magnetic Resonance Imaging as a Marker for Parkinson's Disease. Mov. Disord. 2017, 32, 619–623.

[5]. Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes DiseasePrediction Using Data Mining", 2017 International Conference on Innovationsin Information, Embedded and Communication Systems (ICIIECS)

[6]. Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, "Implementationof a Web Application to Predict Diabetes Disease: An Approach Using MachineLearning Algorithm", 2018 21st International Conference of Computer andInformation Technology (ICCIT)

[7]. Dickson, D.W. Neuropathology of Parkinson disease. Parkinsonism Relat. Disord. 2018, 46 (Suppl. 1), S30–S33.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, April 2023

[8]. Priyanka Sonar, Prof. K. JayaMalini, "Diabetes Prediction Using DifferentMachine Learning Approaches", Proceedings of the Third InternationalConference on Computing Methodologies and Communication (ICCMC 2019)IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4

[9]. International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium: International Diabetes Federation, 2019.

[10]. American Diabetes Association. Standards of medical care in diabetes-2020. Diabetes Care 2020; 43(Suppl. 1): S14–S31.

[11]. Trends in coronary Heart Disease EpidemiologyCenter for Disease Control and Prevention (Heart Disease Facts).

[12]. Asian Pacific Journal of Global Trend of Cancer Mortality rate: A 25-year study.

[13]. International Diabetes Federation: Expenditure and deaths related to diabetes.

[14]. Naveen Kishore G,V .Rajesh ,A.Vamsi Akki Reddy, K.Sumedh,T.rajesh Sai Reddy, "Prediction Of DiabetesUsing Machine Learning Classification Algorithms".

[15]. M.Marimuthu ,S.Deivarani ,R.Gayatri, "Analysis of Heart Disease Prediction using Machine Learning Techniques".

[16]. Purushottam, Richa Sharma ,Dr. Kanak Saxena, "Efficient Heart Disease Prediction System".

[17]. Adil Hussain She, Dr. Pawan Kumar Chaurasia," A Review on Heart Disease Prediction using Machine Learning Techniques".

[18]. Times Of India: Cancer cases upswing 10% in 4 years to 13.9 lakh.

[19]. Epidemiology of Diabetes :A report of Indian Heart Association.

AUTHORS PROFILE

1. Emad Naushad(1902220100066) Student, B.Tech Computer Science & Engineering, I.T.S Engineering College, Greater Noida.
2. Bhavishya Raj(1902220100056) Student, B.Tech Computer Science & Engineering, I.T.S Engineering College, Greater Noida.
3. Arpit Nirvan(1902220100038) Student, B.Tech Computer Science & Engineering, I.T.S Engineering College, Greater Noida.
4. Vrinda Sachdeva Associate Professor of Department of Computer Science & Engineering at I.T.S Engineering College, Greater Noida.

View publication stats