Grenze International Journal of Engineering and Technology, June Issue

A Comparative Analysis Multiple Disease Prediction using Machine Learning Models

Swathi Ch[1] and Ramesh Cheripelli[2]
[1]Assistant Professor, Dept of CSE, G Narayanamma Institute of Technology and Science, Hyderabad, Telanagana, India
Email: chswathi0508@gmail.com
[2]Vidya Jyothi Institute of Technology, Assoicate Professor, Dept of IT, Hyderabad, Telanagana, India
Email: chramesh23@gmail.com

***Abstract*— The diagnosis of diseases at an earlier stage is critically important to the overall improvement of healthcare outcomes because it paves the way for timely interventions and more effective treatment choices. The development of the Smart Health Prediction System as a subfield of medical science has paved the way for reliable disease prediction by utilizing data mining strategies. This was previously impossible. However, the currently available systems have a number of drawbacks, including a slow generation rate for analysis reports, low operational efficiency, high operational costs, and poor accuracy. In this context, the model that is being suggested tries to properly forecast the overall health condition of individuals by leveraging measurable parameters such as TH11, Spo2, pulse rate, and MQ2 parameters. Specifically, the goal of this work is to improve overall health outcomes.**

***Index Terms*— Health Care , Machine Learning , Multiple Disease , Smart Health Prediction.**

I. INTRODUCTION

Your goal is to simulate the usual appearance of papers in a Conference Proceedings or Journal Publications. We are requesting that you follow these guidelines as closely as possible.
The medical industry is always making strides toward better patient care, and recent developments in technology have prepared the way for innovative new approaches to take in this direction. In recent years, machine learning has emerged as a strong tool in a variety of fields, including disease prediction. One of these fields is healthcare. Our research endeavor is centered on the application of machine learning strategies, more especially Support Vector Machines (SVM) and Logistic Regression, in order to forecast the development of three key diseases: diabetes, cardiovascular illness, and Parkinson's disease. The core Indian dataset that can be accessed in the UCI repository is going to serve as the basis for our investigation. This dataset includes useful information pertaining to age, gender, and symptoms that are associated with the diseases that were previously discussed. We hope to be able to assess the likelihood that patients may get these diseases during the next five years by conducting research on the dataset at hand and making use of machine learning methods.

A. Motivation

Our effort was inspired by the need to modernize and automate the healthcare industry, particularly illness prevention and early detection. Modern healthcare is reactive, meaning interventions are made after the disease

Grenze ID: 01.GIJET.10.2.548

has worsened. This technique strains healthcare resources, raises costs, and harms patient outcomes. We seek to improve patient outcomes, reduce healthcare costs, and create a more sustainable healthcare system by shifting from reactive to proactive care. Our project uses machine learning and data analysis to pro- duce a credible illness prediction system for complicated conditions including diabetes, heart disease, and Parkinson's disease. certain new technologies allow us to detect persons at risk of acquiring certain conditions before symptoms appear. Early detection allows medical professionals to provide timely and targeted interventions, therapies, and lifestyle adjustments.

Our study is driven by the need to improve health care by preventing sickness and detecting symptoms early. We want to create a prediction system that uses machine learn- ing and data analysis to enhance patient outcomes, prompt early intervention, and reduce healthcare costs. Our initiative's advantages to society could lead to a healthier future for individuals, communities, and society. These benefits include better public health, evidence-based decision-making, and medical research discoveries.

II. Related Work

A. Analyzing the Diabetes Literature

Decision Tree A decision tree is a classifier with a tree structure, where each node represents a decision and each branch represents a result. It starts with a root node and then branches off at the dominating input feature. This procedure is repeated until every input has been placed, and the final node stores the weights on the basis of which the input is categorized. The fields of law and business, as well as civil planning, finance, engineering, risk management, etc., are only a few of the many real-world applications of the decision tree algorithm.

Random Forest The Random Forest method is used for supervised learning. The forest is constructed using a collection of decision trees. Due to its ability to generate and merge several decision trees, Random Forest is able to provide a more reliable and consistent prediction. The higher the number of trees in the forest, the greater the accuracy. Because of its flexibility and ease of implementation, it has become a popular machine learning algorithm. It's capable of carrying out classification and regression analyses. Banking, healthcare, e- commerce, and the stock market are just few of the many potential uses.

Naive Bayes The incorrect premise that all data points may be evaluated separately underlies Naive Bayes classification. Thus, one trait is unrelated to all others. Classifying texts is its main use. Theorem, Rule, or Law is based on conditional probability.Naive Bayes classifiers have five steps. First classify, then summarize data. Classification summary follows. Calculate class probabilities using Bayes. Reach the expected conclusion.

Three data sets are used to evaluate the ML model. The first two are from Nagpur's Lata Mangeshkar Hospital and N.K.P. Salve Institute of Medical Science and Research Centre; the third is from Mendeley Data. A database was created using patient files to create the diabetes dataset. The data includes medical and lab tests. The dataset includes diabetes factors including gender, age, BMI, HDL, LDL, TG, cholesterol, HbA1c, and TG. It also shows TG, cholesterol, HbA1c, HDL, LDL, and HDL. There are 1313 samples: 60 from Type-1 diabetics, 1150 from Type-2 diabetics, and 103 from non-diabetics.

B. Literature Review on Heart Disease

Using the same inputs as our work, the Naive Bayes, Support Vector Machine, and Func- tional Trees methods were able to predict the presence of cardiac illnesses with an accuracy of 84.5%.

Using the same data, a solo use of Nave Bayes achieved an accuracy of 86.4%. With the same dataset as this study, another system used a variety of algorithms to predict the presence of heart disease: Logistic Regression, KNN, NN, SVM, NB, Decision Tree, and RF, as well as three feature selection techniques: Relief, mRMR, and LASSO. Predictions made using the Logistic Regression technique achieved an accuracy of 89%. Using data on patients' primary health characteristics, this paper proposes multiple machine learning algorithms for predicting cardiovascular illnesses. Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB) were shown to be effective classification techniques for creating prediction models in this paper. Before the models were built, they underwent data preprocessing and feature selection. Accuracy, precision, recall, and the F1-score were used to rank the models. The SVM model was the most accurate with a 91.67% success rate.

C. Literature Review on Parkinson's Disease

In this research, we used Naive Bayes to foretell how well the dataset would perform. The data was explored, statistically analyzed, and mined using Rapid miner 7.6.001. Naive Bayes achieves an accuracy of 98.5% and a precision of 99.75% in its predictions.

Random Forest is compared to Gradient Boosted Trees and utilized for classification in this system. With an accuracy of up to 86.4%, these results are being separated into PSP, De Novo Parkinson's Disease, and Stable Parkinson's Disease, while Gradient Boosted Trees could only achieve 70% accuracy. Furthermore, compared to Gradient Boosted Trees, where the highest accuracy achieved was just 85%, Random Forest achieved a maximum of 90%. Two machine learning methods, logistic regression and support vector machine, are used to make predictions about Parkinson's disease. The findings demonstrate that the Regression model is more successful than the support vector machine in terms of accuracy. Features were selected with care to target the root cause of the issue rather than just the symptoms. On the basis of four days of monitoring with two PD subjects, we present encouraging preliminary results (106 International Journal for Modern Trends in Science and Technology).

III. Proposed System

Our suggested approach will leverage machine learning and other data analysis methods to change disease prediction. Precision and early diagnosis of diabetes, heart disease, and Parkinson's disease are in demand. This need is growing as healthcare data becomes more accessible. We use complex algorithms, exhaustive data analysis, and rigorous model eval- uation to give doctors a powerful tool for proactive sickness management and intervention. The system speeds data collection, preprocessing, feature extraction, and model validation for accurate disease forecasts.

A. Proposed System Architecture

Our illness prediction system relies on several key components to accurately predict dis- eases. From data collection to model evaluation, each module is crucial to the workflow. Let's examine each module's parts:

1. Data collection: Our disease prediction system relies on data gathering. We will carefully choose and collect disease prediction data in this part. We use JM1 software data to finish our job. This dataset underpins our predictive model creation and evaluation. For accurate and reliable prediction models, comprehensive, high-quality data is essential.

2. Data Preprocessing: The data preprocessing module organises data for analysis. This module involves several key actions to ensure data integrity, consistency, and com- patibility. The first step is data formatting, which prepares the data for analysis. This may require data to be exported from the relational database and saved as a flat file or converted to a structure. Next, clean the data by removing duplicates, outliers, and miss- ing numbers. This approach reduces dataset noise, preserving its integrity. Finally, data sampling can lower the amount of a dataset, making it easier to handle and analyze while maintaining representativeness.

3. Extraction of Features: Feature extraction is crucial to finding and selecting the most relevant qualities from preprocessed data. The approach relies on feature extraction. Here, the original properties are changed into new features that better capture illness prediction information. Dimensionality reduction and transformation are used to extract meaningful and informative features from a dataset. This approach reduces computational complexity and increases prediction models' capacity to capture data patterns and correlations.

4. Evaluation Model: The evaluation model module tests disease prediction models. This step divides preprocessed data and features into training and testing sets. The training set trains prediction models, whereas the testing set evaluates their accuracy. We used Random Forest classifiers for our project since they work well for text categorization. Accuracy, which indicates the percentage of correct test data predictions, is used to evaluate model performance. The evaluation provides insights into the models' predictive capabilities and a basis for examining their performance in real-world sickness prediction scenarios.

Our disease prediction system is able to create reliable predictions by adhering to the architecture depicted in Fig.1. This architecture allows our system to make use of the power of data gathering, preprocessing, feature extraction, and model evaluation. Each module contributes significantly to the process of honing and analyzing the data, selecting informative characteristics, and evaluating the effectiveness of the models. This integrated approach enables medical practitioners to make informed decisions, which in turn helps facilitate the early detection of disease, enhances patient care, and ultimately improves patient outcomes.

IV. Methodology

Several important steps must be taken during the implementation and testing part of the current system to make sure that accurate disease prediction is achieved. It starts with choosing and getting ready the right datasets. Next, the gear and software needs are spelled out. The system is then split up into modules, and each module is in charge of a different set of tasks, such as collecting data, preparing it, extracting features, and evaluating it. Training and

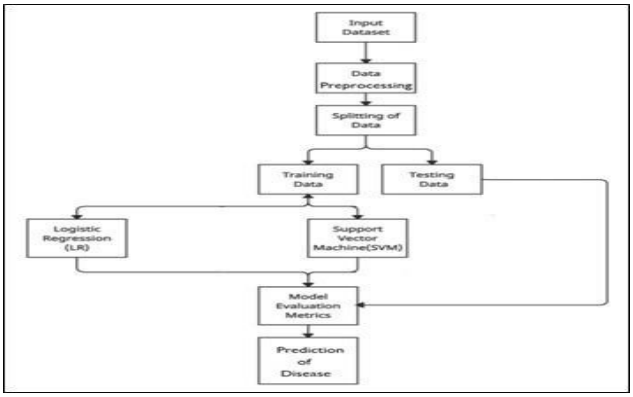

Fig. 1. System Architecture

testing sets are used in strict testing processes, and different evaluation metrics are used to judge how well the model works. The system is improved by carefully putting it into use and trying it so that it can make reliable and accurate predictions about diseases.

A. Brief Description of Datasets

The information we used in our project came from Yashoda Hospital, a well-known Indian hospital. These sets of data contain details about diabetes, heart disease, and Parkinson's disease, which are all very common illnesses. This information comes from carefully col- lecting information from real people, so it is relevant and useful for the Indian healthcare system. Each dataset has a number of characteristics that tell us important things about the patients' health and how the diseases might progress. We want to use these datasets to build strong machine learning models that can correctly predict how likely it is that Indian patients will get these diseases. This will allow for earlier detection, treatment, and better health outcomes.

B. Diabetes Database

High blood sugar is a sign of diabetes, a long-term metabolic disease. The diabetes dataset we used for our project has important features that help us guess if a patient has diabetes. Some of these characteristics are blood sugar, blood pressure, body mass index (BMI), insulin levels, age, and the number of births. Glucose levels show how well someone is controlling their blood sugar, while blood pressure and body mass index (BMI) show how healthy they are generally. Types of diabetes risk factors can also be found by looking at things like insulin levels and the number of births. By looking at these factors, our machine learning models hope to correctly guess how likely it is that a person will develop diabetes. This will allow for early diagnosis and personalized care plans for each patient. The following factors were taken into account for the diabetes dataset:

1. Pregnancies*:* How many times you've been pregnant.
2. Glucose: The amount of glucose in the blood after a 2-hour oral glucose tolerance test.
3. Blood Pressure: The diastolic blood pressure will be shown.
4. The thickness of the muscle on the triceps is measured in millimeters.
5. Insulin: 2-Hour serum insulin (mu U/ml).
6. Body mass index, or BMI, is the ratio of a person's weight in kilograms to their height in meters.
7. The diabetes pedigree function shows how genes affect the chance of getting diabetes.
8. Age: How old you are in years.
9. The result is the goal variable that tells us if a person has diabetes (0 = No, 1 = Yes). Figure 2 shows a small part of the diabetes information that was looked at for this project.

C. Dataset for Heart Disease

Heart disease is a common health problem that includes a number of different heart prob- lems. The heart disease information we used for our project is made up of important factors that help us guess if a patient has heart disease. Some of these factors are age, gender, type of chest pain, resting blood pressure, serum cholesterol level, fasting blood sugar, electrocardiographic results, highest heart rate reached, exercise- induced angina, and having thalassemia as a blood disease. These factors give us important information about the patient's age, gender, cardiovascular health markers, and exercise-induced re- actions. This helps our machine learning models

accurately guess if the patient has heart disease. By looking at these factors, we hope to improve early detection and give people more targeted care. The following characteristics were taken into account for the heart disease dataset:

1. Age: How old the person is.
2. the patient's sex (0 = female, 1 = male).
3. Type of chest pain (1 = normal angina, 2 = atypical angina, 3 = pain that isn't related to angina, 4 = no symptoms).
4. Blood pressure at rest (mm Hg) was measured. This is the amount of cholesterol in the blood (mg/dl).
5. When you wake up, your blood sugar is higher than 120 mg/dl (0 = False, 1 = True).
6. Thalach: Heart rate reached its highest point.
7. Exang: Exercise-induced angina (0 = No, 1 = Yes).
8. In Oldpeak, activity causes ST depression compared to rest.
9. means the peak exercise ST section is going up, 2 means it's flat, and 3 means it's going down.
10. Ca: The number of large blood veins with a color (0–3).
11. Thalassemia is a blood disease (3 = Normal, 6 = Fixed defect, 7 = Reversible defect).
12. The measure that shows if someone has heart disease (0 = No, 1 = Yes).

Figure 3 shows a small part of the heart disease information that was looked at for this project.

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
3	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
2	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
1	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
5	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1

Fig. 2. Diabetes Dataset Sample

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1

Fig. 3. Heart Disease Dataset Sample

D. Dataset for Parkinson's Disease

Parkinson's disease is a neurodegenerative disease that affects the nervous system. It is mostly known for its motor signs, like tremors, stiffness, and trouble moving. The infor- mation we used for our project on Parkinson's disease has important features that help us predict this condition. There are measurements for the basic frequency, jitter, shimmer, changes in amplitude, noise-to-harmonic ratio, fractal scaling exponent, and nonlinear dynamic complexity among these qualities. Our machine learning models can correctly predict if someone has Parkinson's disease based on these attributes, which tell us a lot about their voice and their motor skills. Our goal is to make early detection, personalized care, and better quality of life for patients easier by looking at these factors. The following characteristics were taken into account for the heart disease dataset:

1. The fundamental frequency (Hz) is given by MDVP:Fo(Hz).
2. The highest frequency component (Hz) is MDVP:Fhi(Hz). Third, MDVP:Flo(Hz) is the lowest frequency component (Hz).
3. MDVP:Jitter(%): This variable measures changes in the length of the basic frequency cycle.
4. MDVP:Jitter(Abs): Finds the absolute difference between two or more times.
5. MDVP:RAP: Finds the relative amplitude change.
6. The MDVP:PPQ measures changes that happen at a frequency of 1 Hz to 4 Hz.
7. Jitter: DDP: Finds the average absolute difference of the changes between periods.

8. MDVP:Shimmer: Checks for changes in intensity.
9. MDVP:Shimmer(dB): Finds the change in volume in decibels.
10. Shimmer:APQ3: Checks how the intensity changes in the first 3 milliseconds.
11. Shimmer:APQ5: Checks how the intensity changes in the first 5 milliseconds.
12. MDVP:APQ: Checks how the amplitude changes between times that are next to each other.
13. Shimmer:DDA: Finds changes in amplitude by comparing the absolute differences between times.
14. NHR: Finds the ratio of noise to harmonics.
15. HNR measures the ratio of sound waves that repeat to those that don't.
16. Status: The number that tells us if someone has Parkinson's disease (0 = Healthy, 1 = Parkinson's disease).
17. RPDE stands for recurrence period density entropy.
18. DFA: Exponent for signal fractal scaling.
19. Spread1, Spread2, and D2 are nonlinear measures of dynamic complexity. PPE stands for "perceptual pitch entropy."

Fig 5 shows a small part of the Parkinson's disease information that was looked at for this project.

MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
0.426	0.02182	0.0313	0.02971	0.06545	0.02	21.033	1	0.414783	0.82	-4.81303	0.266482	2.3	0.28
0.626	0.03134	0.04518	0.04368	0.09403	0.02	19.085	1	0.458359	0.82	-4.07519	0.33559	2.49	0.37
0.482	0.02757	0.03858	0.0359	0.0827	0.01	20.651	1	0.429895	0.83	-4.44318	0.311173	2.34	0.33
0.517	0.02924	0.04005	0.03772	0.08771	0.01	20.644	1	0.434969	0.82	-4.1175	0.334147	2.41	0.37
0.584	0.0349	0.04825	0.04465	0.1047	0.02	19.649	1	0.417356	0.82	-3.74779	0.234513	2.33	0.41
0.456	0.02328	0.03526	0.03243	0.06985	0.01	21.378	1	0.415564	0.83	-4.24287	0.299111	2.19	0.36
0.14	0.00779	0.00937	0.01351	0.02337	0.01	24.886	1	0.59604	0.76	-5.63432	0.257682	1.85	0.21
0.134	0.00829	0.00946	0.01256	0.02487	0	26.892	1	0.63742	0.76	-6.1676	0.183721	2.06	0.16
0.191	0.01073	0.01277	0.01717	0.03218	0.01	21.812	1	0.615551	0.77	-5.49868	0.327769	2.32	0.23

Fig. 5. Parkinson's disease dataset sample

V. RESULTS

Our Multiple Disease Prediction System's user interface (UI) is shown in the Results section to make forecasting diabetes, heart disease, and Parkinson's disease easy. The engaging and easy Streamlit online software lets users enter their glucose, blood pressure, and other vital health data. The system then processes this data and uses complex machine learning methods like Support Vector Machines and Logistic Regression to make real-time illness category predictions. Users can utilize the attractive UI to make health decisions and take preventative measures based on forecast results. Our technology empowers people to manage their health and live healthier with its user-friendly design and accurate disease prediction.

A. Result for Diabetes

The Diabetes Prediction section shows our Multiple Disease Prediction System's diabetes predictions. The positive test cases in Fig. 6 (a) indicate that the system predicts diabetes based on attribute values. These forecasts inspire people to seek medical assistance and adjust their lifestyles to control diabetes. The negative test cases as shown in Fig 6 (b).

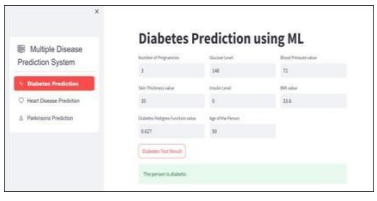

Fig. 6. (a) Positive Test Cases (b) Negative Test Cases*B. Results for Heart Disease*

The Heart Disease Prediction findings are based on a sophisticated machine learning tech- nique that evaluates the provided attribute values to calculate the likelihood of an indi- vidual having heart disease. The positive test cases as shown in Fig 7 (a) illustrate instances when the Heart Disease Prediction System shows a higher chance of heart disease based on the specified qualities. The approach uses essential parameters such as age, sex, chest pain kinds, resting blood pressure, serum cholestoral level, and other clinical markers to identify patients at risk. For individuals with favorable predictions, it is vital to consult with healthcare professionals for early diagnosis and individualized treatment regimens. Timely intervention can considerably improve the prognosis and assist manage

cardiac disease effectively, boosting the overall quality of life for affected persons. In contrast, Fig.7(b) shows negative test scenarios.

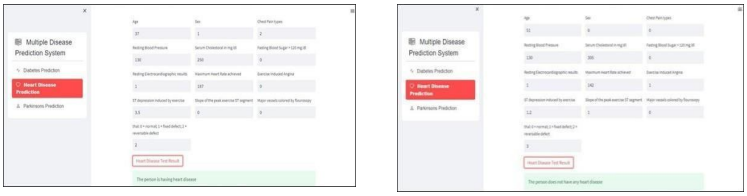

Fig. 7. (a) Positive Test Cases (b) Negative Test Cases

C. Parkinson's results

Advanced machine learning algorithm examines attribute values to predict Parkinson's illness. Fig. 8(a) shows positive test scenarios where the Parkinson's Disease Prediction System forecasts a greater risk based on features. To determine risk, voice measurements, tremor markers, and other clinical aspects are thoroughly evaluated. Positive predictions require neurologists or movement problem specialists for full evaluations and customized treatment approaches. Early identification and treatment of Parkinson's disease can en- hance outcomes and quality of life for patients and their families.In contrast, Fig.8(b) shows negative test scenarios.

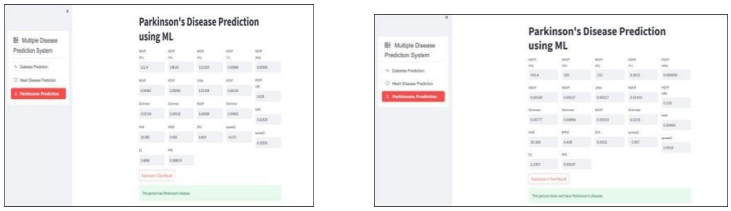

Fig. 8. (a) Positive Test Cases (b) Negative Test Cases

VI. Conclusion And Future Enhancements

Our Multiple Disease Prediction System uses cutting-edge machine learning algorithms, advanced data analysis, and easy user interfaces to forecast diabetes, heart disease, and Parkinson's disease accurately. We use Support Vector Machines and Logistic Regression to identify complex data patterns and correlations to create predictions. Our easy-to-useinterface lets users enter attribute values and get real-time projections for proactive health management. The system's reliable assessments help identify and treat conditions early, improving health and quality of life. As our system evolves and expands, it could alter disease prediction and preventative treatment, improving public health worldwide. We want to create a mobile app so consumers may access the system from their cellphones and get health insights quickly.

References

[1] F. Otoom, et al "Effective Diagnosis and Monitoring of Heart Disease," International Journal ofSoftware Engineering and its Applications (IJSEA), vol. 9, no. 1, pp. 143–156, 2015.

[2] U. Haq, et al "A Hybrid Intelligent System Framework for the Prediction of Heart Disease usingMachine Learning Algorithms," Mobile Information Systems, 2018.

[3] Carlo Ricciardi, et al, "Using Gait Analysis' Parameters to Classify Parkinsonism: A Data Mining Approach", Machine Learning and - An International Journal, 2020.

[4] Chaimaa Boukhatem, et al "Heart Disease Prediction Using Machine Learning," (ASET), 2022.

[5] Gokila et al, "Prediction of Parkinson's Diseases using SVM and Logistic Regression Algorithm," ,International Journal for Modern Science and Technology, 2021.

[6] K. Vembandasamyp, R. R. Sasipriyap, and E. Deepap, "Heart Diseases Detection Using Naive Bayes Algorithm," International Journal of Innovative Science Engineering and Technology (IJISET), vol. 2, no. 9, 2015.

[7] Nidhi Kosarkar, et al "Disease Prediction using Machine Learning," International Conference on Emerg-ing Trends in Engineering & Technology Signal and Information Processing (ICETET-SIP), 2022.

[8] Rahul R et al "Predictive Model for Parkinson's Disease through Naive Bayes Classification," Interna- tional Research Journal of Engineering vol. 07, issue: 10, 2020.

[9] Yogita Dubey, et al "Disease Prediction and Classification using Machine Learning Algorithms," In- ternational Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), 2021.

[10] Subramanian, et al. (2017). "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles." Cell, 171(6): 1437-1452.e17.

[11] Chen Y.et al. (2016) “Gene expression inference with deep learning. Bioinformatics,” 32(12): 1832- 1839.
[12] Li W, et al. (2019) “Gene Expression Value Prediction Based on XGBoost Algorithm.” Front. Genet.,10: 1077.
[13] Clayman C.L etal (2020) “K-means clustering and principal components analysis of microarray dataof L1000 Landmark Genes.” Procedia Computer Science,168: 97-104.
[14] Libbrecht M.W., Noble W.S. (2015) “Machine learning in genetics and genomics.” Nat Rev Genet, 16(6): 321-332.
[15] Xiao Y., Wu J., Lin Z., et al. (2018) “A deep learning-based multi-model ensemble method for cancer prediction.” Computer Methods and Programs , 153: 1-9.
[16] Liu P.et al. (2019) “Diagnosis of T-cell–mediated kidney ejection in formalin-fixed, paraffin-embeddedtissues using RNA-Seq–based machine learning algorithms.” Human Pathology, 84, 283-290.
[17] Cheripelli, R New Challenges and its Security, Privacy Aspects on Blockchain Systems,14th Inter- national Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023,pages 1491-1497.
[18] Tabares-Soto et al. (2020) “A comparative study of machine learning and deep learning algorithmsto classify cancer types based on microarray gene expression data.” PeerJ Comput. Sci., 6: e270. DOI 10.7717/peerjcs.270.
[19] Alanni R.et al. (2019) “A novel gene selection algorithm for cancer classification using microarray datasets.” BMC Medical Genomics, 12:10.
[20] Swathi, C. and Cheripelli, R,Computational Learning Model for Prediction of Parkinson Disease Using Machine Learning,Smart Innovation, Systems and Technologies,2023 volume 313,pages 331-341.
[21] Chen, T.Guestrin, C. (2016) “XGBoost: A Scalable Tree Boosting System.” Proceedings of the KDD’, 16, 1-10.
[22] Ma S., and Dai Y. (2011) “Principal component analysis based methods in bioinformatics studies.” Briefings In Bioinformatics, 12(6): 714-722.
[23] Sáez J.A., M. (2016) “Analyzing the oversampling of different classes and types of examples in multi- class imbalanced datasets.” Pattern Recognition, 57, 164-178.
[24] Ramesh ,Comparative analysis of applications of identity-based cryptosystem in IoT,Electronic Gov- ernment,2017 Volume13,pages 314-323
[25] Enache et al. (2019) “The GCTx format and cmap Py, R, M, J packages: resources for optimizedstorage and integrated traversal of annotated dense matrices.” Bioinformatics, 35 (8): 1427-1429.
[26] Buscher, et al. (2017) “Natural variation of macrophage activation as disease-relevant phenotype pre- dictive of inflammation and cancer survival.” 16041.
[27] Cheripelli, R. New Model to Store and Manage Private Healthcare Records Se- curely Using Block Chain Technologies. In: Islam, A.K.M.M., Uddin, (eds) Banga- bandhu and Digital Bangladesh. ICBBDB 2021. Communications in Computer and Information Science, vol 1550. Springer, Cham.