”Health Predictor: A Flask Web Application for Depression and Cardiovascular Disease Prediction”

MSc Research Project
Artificial Intelligence

Vamshi Goud Gouni
Student ID: 22122290

School of Computing
National College of Ireland

Supervisor: Rejwanul Haque

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Vamshi Goud Gouni
Student ID:	22122290
Programme:	Artificial Intelligence
Year:	2023
Module:	MSc Research Project
Supervisor:	Rejwanul Haque
Submission Due Date:	31/01/2024
Project Title:	"Health Predictor: A Flask Web Application for Depression and Cardiovascular Disease Prediction"
Word Count:	7919
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Vamshi Goud Gouni
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

”Health Predictor: A Flask Web Application for Depression and Cardiovascular Disease Prediction”

Vamshi Goud Gouni
22122290

Abstract

The world’s problem of depression and cardiovascular illnesses is addressed by using the integration of Machine learning in the Flask online application Health Predictor. The focus on early intervention and tailored medication aligns with the growing demand for accessible health solutions. The main goal is to incorporate precise machine learning models to anticipate depression and cardiovascular within a user-friendly Flask application. Some specific goals include developing models using appropriate datasets, integrating apps, and evaluating performance. Some algorithms also use kaggle datasets for training. Cardiovascular disease prediction showed 72 percent accuracy for logistic regression, which proved resilient. An ensemble method, Voting Classifier (Random Forest and Gradient Boosting) with outstanding 73 percent accuracy shows the relevance of model selection. In depression prediction, Random Forest scored higher than Logistic Regression with an accuracy of 83.22 percent. While some difficulties were encountered by the ensemble method and resulted in an accuracy of 83.22 percent, these results show the limitation in optimization. The findings in both forecasts showed trade-offs in accuracy, recall, and precision. Some insights include the importance of the ensemble approach, the need for hyperparameter tuning, and achieving the right balance needed for accurate forecasting. The Flask app successfully combines multiple models for ease of application in health assessments.

1 Introduction

The combination of machine learning and healthcare has enabled novel methods of predicting and averting serious health illnesses. The problem of cardiovascular disease is an old one that persists, while the incidence of mental disorders such as depression is increasing today. The “Health Predictor” is our project which will solve this problem using a user-friendly platform to assess health via the incorporation of algorithms predicting cardiovascular diseases and deprivation in a complete Flask web application.

There are cardiovascular and mind disorders that heavily affect the world’s wellness therefore they warrant preventive and easily available care.Lamiae et al. (2022) state that the inclusion of machine learning and smart home technologies can greatly add up to sustainable health care by providing early detection and unique insights. Depression is one of the most prevalent mental health issues. There is a lot of research regarding it. As highlighted by Shameer et al. (2018) and Gao et al. (2018), the use of machine learning for predicting depression can increase the accuracy of diagnosis, and tailor treatment.

On the contrary, cardiovascular diseases remain the leading cause of the global burden of disease (*C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction* (2012). As per Alty et al. (2004), predictive models have been proven useful in spotting those at risk of becoming abusers or abused. Our project gives users the entire health prediction experience by integrating these insights into a Flask web application. Firdaus et al. (2018), and Jain et al. (2019) have noted the significance of machine learning for health, and this underscores the transformative potential of Health Predictor application.

The Health Predictor project seeks to meet the surging demand for complete and simple health prediction tools. By leveraging technology in the management of health, more people have opted to seek help through our Flask web application to check on their cardiovascular risks as well as mental wellness. As observed by Elayan et al. (2021), the targets of our application coincide with the sustainability of healthcare data analysis using IoT-based systems. Our system facilitates forward-looking good health and predictive medicine via presages and personalized knowledge.

However, the project becomes even more significant due to the deficiencies related to the usual healthcare paradigms. Patel et al.(2016) argue that machine learning can help in the investigation of depression using an imaging approach and it shows the dynamic nature of healthcare research. However, for the innovations to have a widespread impact, they must be incorporated into easy-to-use apps. Our Health Predictor fills this gap by offering actionable predictions in an understandable way to people with different technical backgrounds.

The key concern on which the Health Predictor project is centered around involves incorporating machine learning techniques into a Flask-based web application for predicting depression and cardiovascular disease. Some of the important aims of this research involve. The project will, in the first place, endeavor to create accurate models for predictions using related databases relating to cardiovascular and depressive disorders. The models are necessary for the application to produce accurate and reliable health forecasts.

Importantly, the latter phase of the project entails seamlessly unifying both prediction models in an intuitive Flask web app. It highlights making an interface that is not only user friendly but also navigable making it easy for users irrespective of their familiarity with technical issues.

Eventually, the project entails a thorough assessment of the application's accuracy in forecasting health issues. This appraisal will evaluate the efficiency of integrated machine learning models in the Flask web application. The Health Predictor project is targeting to develop a strong and user-oriented system that uses machine learning to make accurate health predictions and thus support healthcare technology and accessibility advancement.

To accomplish these goals, we utilize machine learning algorithms that were trained by us on Kaggle datasets related to cardiovascular disease and depression.

The recognition of the limits associated with the assumptions and predictive model used while developing the Health Predictor app is important. As noted by McMahon et al (2011), the quality of our models is largely dependent on the quality of the input data. The healthcare industry's nature makes the forecasts inherently uncertain. Additionally, the framework design assumes user engagement and data security as per Chekroud, et al.'s (2016) criteria for transference of treatment outcomes across trials.

Therefore, the concluding point of the Health Predictor is about the innovative Flask web application having ML forecasts of cardiovascular and depressive disorders. It is another innovation in the fast-changing field of healthcare technology that deals with the significance of cardiovascular risk and mental health and offers users an active way to cope with one's wellness status. This report will discuss the technical aspects of the Health Predictor application, its development process, and its evaluation in its later sections.

2 Related Work

Artificial intelligence (AI) and machine learning (ML) are transforming medicine and ensuring sustainability in the health sector by applying creative approaches to patients' care. The sustainability of healthcare could be enhanced by integrating machine learning with smart homes as outlined in Lamiae et al. (2022). In addition, their review, released at the International Conference on Smart City Applications, indicates the possible use of AI for health cost savings, remote monitoring of patients, and personalized healthcare.Firdaus et al. (2018) consider the comparison between machine learning and meta-heuristic optimization techniques in intelligent and sustainable healthcare. In the case of long-term healthcare applications, the suitability of different AI approaches is critical and is highlighted in publications such as the African Journal of Computer and ICT. Elayan et al. (2022) use deep federated learning to examine the sustainability of data analytics for healthcare in IoT-based systems. AI's contributions to more effective healthcare data exchange and analysis supporting to long -term health care systems have been investigated in their study that appeared in the IEEE Internet of Things Journal. In their presentation at the 6th International Conference on computing for Sustainable Global Development, Jain and Kaushal (2018) compare different machine learning algorithms for health care. Their study reviews multiple AI techniques providing insights on how they can be employed in the medical field to enhance patient outcomes, reduce expenditures, and develop stronger healthcare systems.

2.1 Machine Learning in Cardiovascular Applications

The application of machine learning techniques in cardiovascular medicine has significantly increased. In a meta-analysis by Krittanawong et al. (2020) a detailed account of the application of machine learning in the prediction of cardiovascular diseases is given. Early intervention would help reduce demand at health facilities and this paper emphasizes the role played by AI-based models in the prediction of cardiovascular disease risk. Dinesh et al. (2018) emphasize predicting cardiovascular diseases which can be diagnosed early thereby preventing. Using the practical application of machine learning in health care settings", is the title of the presentation they gave at the 2018 International Conference on Current Trends towards Converging Technologies. Just like that, Shameer et al. (2018) examines the development of machine learning in the field of cardiovascular medicine. The evaluation evaluates where AI applications currently stand, highlighting challenges and prospects. This underscores the need for machine learning techniques to further advance to adequately support the management of cardiovascular disease. In addition, the growing number of studies emphasized on use of machine learning in cardiology towards personalized therapies, early diagnosis, and early prediction to improve sustainability in the field.

2.2 Machine Learning in Depression Prediction

Machine learning offers alternative ways of understanding, labeling, and predicting depression, one of the leading mental disorders.Gao et al. (2018) considered machine learning for severe depression including classification, and treatment outcome prediction. AI helps to improve our understanding of depression and develop personalized treatment approaches – their study published in CNS Neuroscience and Therapeutics reveals. Machine learning classifiers are applied in a one-of-a-kind setting of older individuals by Firdaus et al. (2018) to predict depression. One of the recent studies on machine learning was presented by Hossain et al., through their work in the International Journal of Computer Applications. Their study reveals that machine learning can also be customized for different demographic groups and could be used in early detection. Patel et al. (2015), apply imaging techniques that use machine learning approaches, in research on depression. Their research, which was published in NeuroImage: Artificial Intelligence in Clinical, reveals the neuronal pathways behind sadness from brain imaging data.

2.3 Advancements in Personalized Cardiovascular Treatment

The use of AI in cardiovascular medicine nowadays has shown how tailoring the treatment regimen for each patient could be done. Krittanawong et al. (2020). A meta-analysis on the application of machine learning predicting cardiovascular diseases. The Scientific Reports. The study stresses the importance of early intervention and positions these AI-driven models as useful tools for assessing the risk of cardiovascular diseases to alleviate the pressure placed on healthcare systems Krittanawong et al. (2020). This aligns with the growing shift toward personalized medicine where artificial intelligence is pivotal in developing tailor-made treatment protocols aimed at improving patient outcomes while enhancing resource utilization in cardiovascular care.

2.4 Early Detection of Depression through Advanced Imaging and AI

Integrating high-end imaging, machine learning has become a powerful tool in the early diagnosis of depression. The work done by Patel et al. (2015) considered the use of machine learning methods for brain imaging when exploring depression research. According to Patel et al. (2015) this research, which was published in NeuroImage: Clinical demonstrates the use of artificial intelligence (AI) in reading brain imaging data and unmasking the neural networks of depression, offering a promising way for earlier detection and prevention. The combination of neuroimaging and machine learning enhances our understanding of molecule levels of depression as well as offers great prospects for early intervention and individual treatment planning. This demonstrates the importance of high tech in resolving mental problems and building stronger mental health systems.

2.5 Challenges and Ethical Considerations in AI-Driven HealthcareI

While artificial intelligence offers numerous benefits in healthcare, it also poses ethical problems such as moral dilemmas.Lamiae et al. (2022) describe the potential of machine learning and smart home technologies as working toward sustainable healthcare. This is compounded by concerns raised by some authorities such as the security of data, confidentiality, appropriate use of patient information, and so on Lamiae et al. (2022)). The credibility of such solutions can only be guaranteed if these challenges are handled directly. Moreover, with the development of these technologies comes the realization of the need for accurate ethical standards and laws, confirming the importance of an effective roll-out approach.

2.6 Cardiovascular Disease Prediction and Intervention

Cardiovascular disease prediction and treatment is one of the critical research topics. The approaches are focused on various machine learning technologies. A study by Collaboration et al. (2012) discusses indicators concerning lipids for forecasting cardiovascular disease. Published as research in JAMA, these indicators have a predictive potential and our further knowledge on cardiovascular risk factors. Amma (2012), provides an example of using computational intelligence in risk assessment in which an algorithm and neural network-based cardiovascular disease prediction system is proposed Amma (2012). Also, Mcmahon et al. (2011) has shown how many biomarkers influence risk assessment as it examines fibrinogen and C-reactive protein as predictors for cardiovascular diseases. *C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction* (2012) performed a study that looked at the relation between systemic lupus erythematosus and its possible effects on cardiovascular diseases. Alty et al. (2004). contribute to the discipline, showing how modern computational methods could be used in risk assessment via the prediction of cardiovascular disease by support vector machines Alty et al. (2004).

2.7 Machine Learning in Depression Treatment Outcomes

Machine learning has immensely increased the prediction of depression treatment results, offering guidance in the individualistic and effective treatment of mental health problems. One research review by Sajjadian et al. (2021), which included a meta-analysis, suggested the utilization of machine learning Priya et al. (2020) use predicting anxiety and sadness in old patients to illustrate machine learning applications' adaptability across demographic groups. According to Chekroud et al. (2016), predictive models should constitute critical elements of the cross-trial prediction strategy used in clinical trials for depression treatment (p. 1). In the paper Usman et al. (2020), an exhaustive review of the research state in the issue has been conducted presenting a broad analysis of the literature on the prediction of depression with various machine learning algorithms. In this regard, Hawes et al. (2023) further our understanding of how machine learning models could be used to predict stress, anxiety, and depression and therefore assess mental health comprehensively using computational approaches.

2.8 Future Directions: Integrating AI into Routine Clinical Practice

The advancement of machine learning in healthcare can lead to a radical change in clinical procedures. Research on the sustainability of healthcare data analysis in IoT-based systems has been highlighted by Elayan et al., (2021). The study shows how AI can lead to the sustainability of healthcare through better data transmission and analyses in healthcare. Elayan et al., 2021, in the IEEE Internet of Things Journal. In addition, utilizing AI in routine clinical work could enhance patient management in general by way of better diagnosis and treatment planning in the future. It is expected that artificial intelligence will be an integral part of future healthcare procedures to increase efficiency, accuracy, and patient outcomes. Ongoing research and breakthroughs indicate this.

3 Methodology

Three main elements make up the research methodology: Depression prediction, cardiovascular disease prediction, and the whole Flask application. In each part, the procedures, supplies, methods for gathering data, and statistical methods are given in detail.

3.1 Cardiovascular Disease Prediction

After obtaining a complete dataset of related health features relevant to Cardiovascular Disease Prediction from Kaggle (cardiovascular disease), the project began. After data collection, stringent techniques of data pre-processing were applied addressing missing values, transforming data, and features normalization

	id	Age	Gender	Height	Weight	Ap_hi	Ap_lo	Cholesterol	glue	smoke	alco	Active	Cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
69995	99993	19240	2	168	76.0	120	80	1	1	1	0	1	0
69996	99995	22601	1	158	126.0	140	90	2	2	0	0	1	1
69997	99996	19066	2	183	105.0	180	90	3	1	0	1	0	1
69998	99998	22431	1	163	72.0	135	80	1	2	0	0	0	1
69999	99999	20540	1	170	72.0	120	80	2	1	0	0	1	0

70000 rows × 13 columns

Figure 1: Cardiovascular Dataset

Following this, EDA was conducted using bar charts, histograms, and a correlation matrix to ascertain how the variables were distributed and related in the dataset.

During the project's effort to build an effective predictive model, the project considered feature selection with the help of both statistical methodologies and domain knowledge to discover essential features. Different machine learning algorithms were used in cardiovascular disease prediction, such as logistic regression, random forest, SVM, etc. The performance of each model was assessed in terms of its accuracy and F1 score. Hyperparameter tuning on best models using grid search cross-validation on top-performing algorithms on accuracy.

The model ensembling approach was used to help improve the predictive power of the model; finally, the best model selected was used to integrate with the Flask Application. In the final stage of the process, a subset of the dataset was used to train selected models whose performance was determined using metrics like accuracy, precision, and recall. Each of these stages was implemented as a holistic approach towards building a strong Cardiovascular Disease Prediction system.

3.2 Depression Prediction

The Depression Prediction project started with buying Kaggle's (depression) depression dataset comprising several mental health and depression-related features. The first stage of data preprocessing consisted of a comprehensive process of transforming and cleaning the dataset to eliminate missing values and homogenize its structure.

	Survey_id	Ville_id	Sex	Age	Married	Number_children	Education_level	Total_members	Gained_asset	Durable_asset	Incoming_salary	Incoming_own_farm	incoming_business	incoming_no_business	incoming_agricultural	Farm_expenses	Labour_primary	Lasting_investment	No_lasting_investman	depressed
0	926	91	1	28	1`	4	10	5	28912201	22861940	0	0	0	0	30028818	31363432	0	28411718	28292707.0	0
1	747	57	1	23	1	3	8	5	28912201	22861940	0	0	0	0	30028818	31363432	0	28411718	28292707.0	1
2	1190	115	1	22	1	3	9	5	28912201	22861940	0	0	0	0	30028818	31363432	0	28411718	28292707.0	0
3	1065	97	1	27	1	2	10	4	52667108	19698904	0	1	0	1	22288055	18751329	0	7781123	69219765.0	0
4	806	42	0	59	0	4	10	6	82606287	17352654	1	0	0	0	53384566	20731006	1	20100562	43419447.0	0
...	...	...	...	...	...	....	....	....	....	...	..	....	....	.....	.....	.....	...	......	....	....
1424	255	22	1	25	1	1	7	5	28912201	22861940	0	0	0	0	30028818	31363432	0	28411718	28292707.0	0
1425	547	69	1	28	1	4	10	6	15711078	24023054	0	1	0	0	23022095	1021536	0	1823477	47384361.0	0
1426	893	184	1	66	0	0	1	1	42440731	22861940	0	1	0	0	12545373	10454478	0	46444572	10454478.0	1
1427	363	75	1	51	1	1	12	5	28912201	22861940	0	0	0	0	30028818	31363432	0	28411718	28292707.0	0
1428	231	12	1	33	0	4	8	5	81678391	22861940	0	1	0	0	20019212	16682677	0	69642126	13012488.0	0

1429 rows × 23 columns

Figure 2: Depression Dataset

Then came the exploratory analysis, which utilized a variety of plots such as bar charts, histograms, and correlation matrices for observing interesting distributions among different variables in the data set.

One of the key elements of the project involved feature engineering that allowed extracting relevant features from textual and categorical data to improve predictive modeling with a better dataset. Machine learning methods such as decision tree algorithms and ensemble approaches proved to be efficient in modeling depression prediction. Then, the two promising models were subjected to stringent hyperparameter tuning, and their predictions were combined using the Voting Classifier ensemble model. In the last stage, depression prediction models were properly evaluated with accuracy, F1 score, and ROC-AUC among the important metrics for comprehensive performance assessment.

The rigorous process included data collection, preprocessing, exploratory analysis, feature engineering, model training, and evaluation all these processes were done comprehensively and systematically. The multilayered methodologies provide a comprehensive understanding of the dataset and allow to develop of robust prediction tools and accurate metrics, which serve as the ground for an effective instrument in the prediction of mental health problems.

3.3 Flask Application

As regards the Flask Application, a careful evaluation of the design specification was done to include only the necessary functionalities for an easy-to-use interface. Simplicity in navigation and rich data visualization capabilities were the major emphasis, aiming at providing the users with a seamless experience. The later stages of implementing solutions involved the utilization of these specifications as the blueprint.

During the Implementation/Solution Development phase, a complete Flask application was developed with both depression and cardiovascular prediction algorithms. It is a dynamic app that allows quick communication with the predictive models and proves to be useful in healthcare information. Carefully, the pickled object for the standard scaler and ensembled model were exported to support the incorporation of machine learning models in the Flask app. The smooth incorporation of these machine learning models into the Flask Application makes it possible for the former to exploit the predictive capabilities of the latter.

Additionally, the application includes a user table with attributes like user ID, username, and user password. Such a database structure allows for incorporating key functions such as user registration and login. The registration and logging in of customers is simple, thereby contributing to the enhanced level of personalization, which enables customers to utilize all the Flask Application features seamlessly.

This is a comprehensive development process that leads to an elegant and presentable Flask application interface. Data transformation and predictions are integrated for ease and provide intuitive visuals as well as health prediction. User-centric designing ensures that individuals, regardless of their technical skills, can easily navigate through the application, hence a powerful and convenient instrument for healthcare professionals as well as health enthusiasts. At its core, the Flask Application is an elegant balance of design perfection, sturdy solution, and customer orientation, which provides a powerful vehicle for health forecasting as well as data management.

The study methodology included extensive data preparation, a careful model selection, and a comprehensive assessment for cardiovascular and depression prediction. It ensured that the platform for getting the prediction results was interactive and simple. These actions together underscore the principles of reproducibility and openness for dependable research results.

4 Design Specification

The Cardiovascular Disease Prediction implementation created multiple machine learning models to achieve maximum prediction accuracy. Among the initialized classifiers were Random Forest, Logistic Regression, Support Vector Machine, Gradient Boosting, k-Nearest Neighbour, Decision Tree, and a Voting Classifier. This latter soft-voting ensemble integrated predictions from Random Forest, Logistic Regression, SVM, Gradient Boosting, k-Nearest Neighbours, and Decision Tree. Specifically, the tuning of hyperparameters was done to enhance the forecast power of Random Forest and Gradient Boosting models. Using the merit of each individual classifier, an improved prediction model was formed with the aid of ensemble technique.

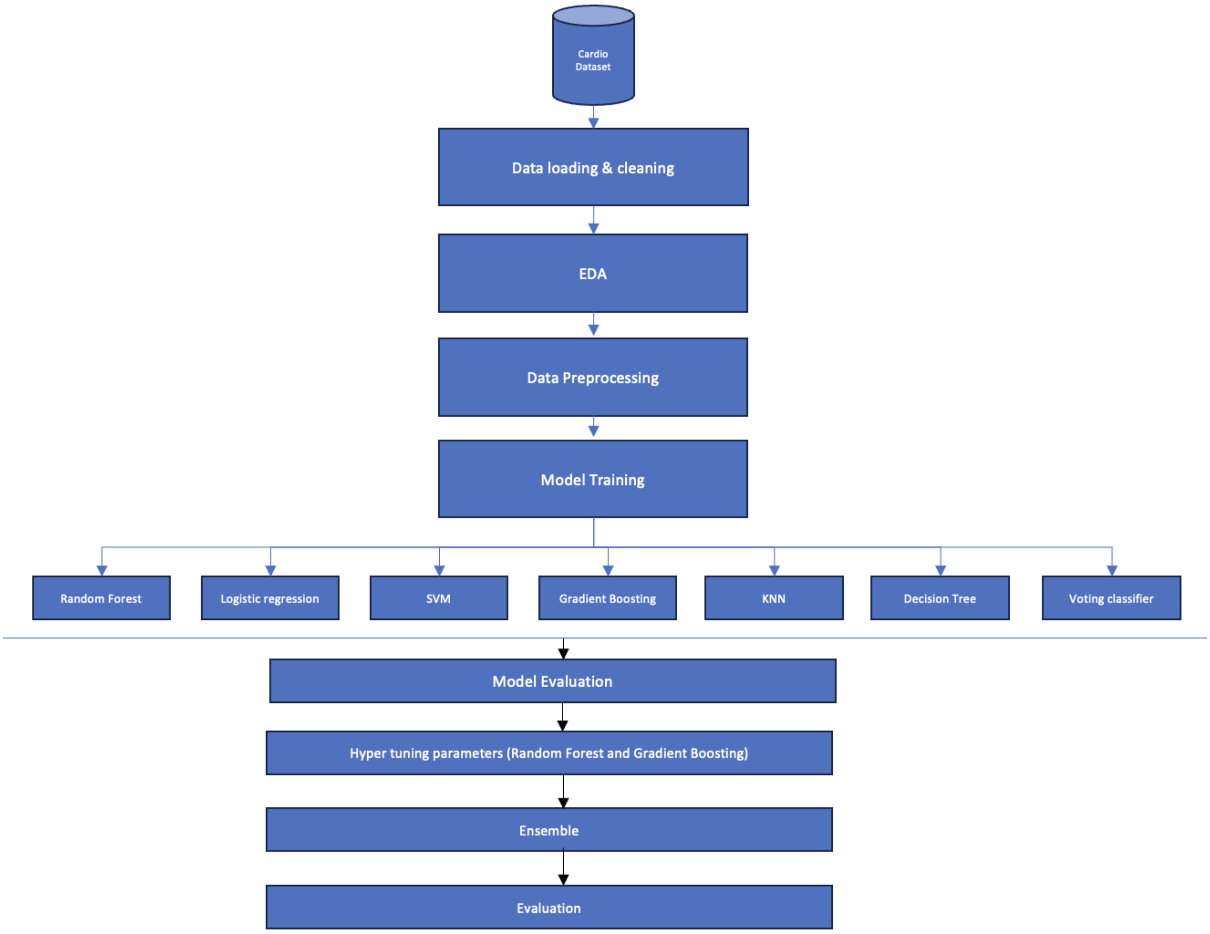

Figure 3: Architecture Diagram on Cardio Dataset

4.1 Architecture diagrams

Similarly, several classifiers, including Random Forest, Logistic Regression, Naive Bayes, SVM, Decision Tree as well as K – Nearest Neighbours, were employed in Depression Prediction. The main objective of hyperparameter tuning was to optimize the parameters of the Random Forest and k-Nearest Neighbours. The resulting customized models were then mixed in a way to improve the overall predictive performance, by leveraging the strengths of each classifier. This strategy took advantage of the complimentary aspects of various algorithms to produce a more reliable and accurate prediction model for depression.

User engagement with predictions considered specifically in the context of the Flask Application combining depression and cardiovascular prediction algorithms. The application imports Pickled Objects for the Ensembled Model and the Standard Scaler smoothly from the Machine Learning Models. Additionally, a "user" database table was created with user-id, username, and password fields to facilitate user function. The implementation of the Register and Login functions became easier with this table, thus enabling a secure and personalized prediction experience for Flask Application users who wanted to learn about depression and cardiovascular indicators.

The above Figures 3 and 4 provide an architecture or visual representation of the coding process for a cardio and depression dataset. They show the systematic procedure or steps in data loading, cleaning, EDA, pre-processing, model training, model evaluation,

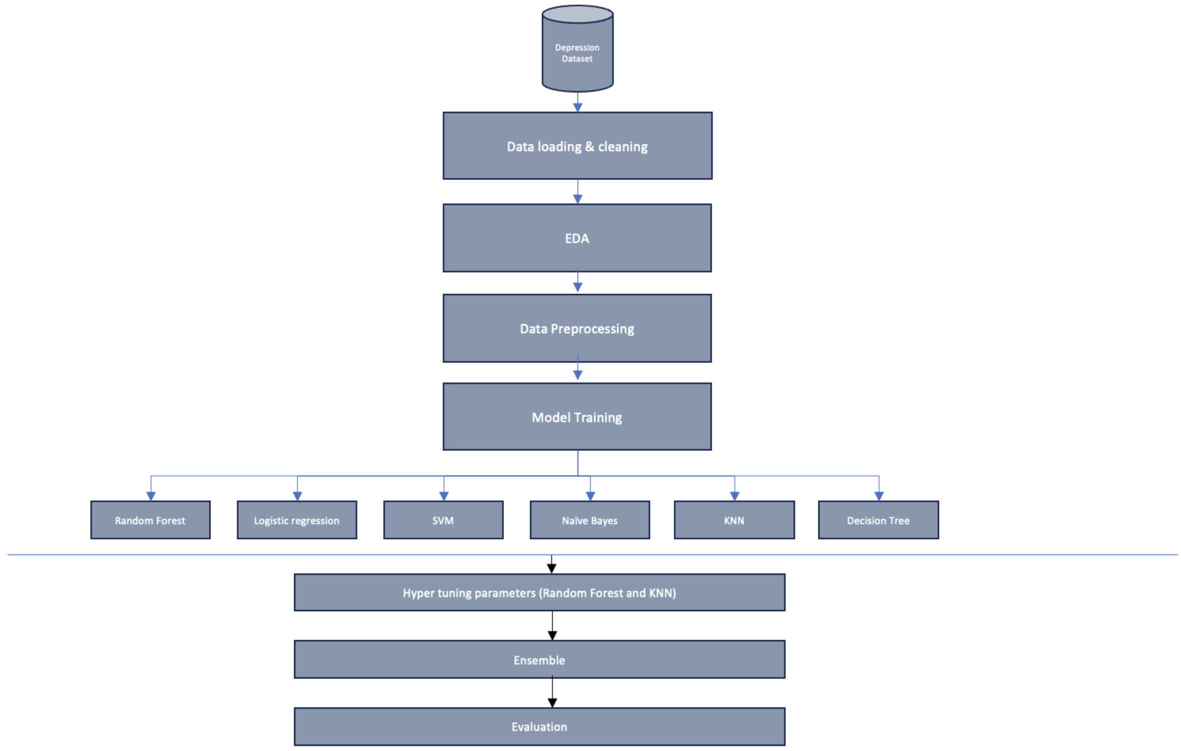

Figure 4: Architecture Diagram on Depression Dataset

hyperparameter tuning, ensemble modelling and final evaluation. In detail, the data loading and cleaning process where the datasets are loaded into the coding platform or environment and performed data cleaning process. The EDA step involves analyzing and visualizing the datasets to gain appropriate insights into their characteristics and structure. The data pre-processing involves activities such as normalization, scaling, encoding or missing value handling to make sure that the data in both datasets are in a suitable format for further analysis. During the model training process, ML models are trained using data cleaning and pre-processing activities. Here, the cardio dataset used trained several models including – Random forest, logistic regression, support vector machine (SVM), gradient boosting, K-nearest neighbours (KNN), decision tree, and voting classifier. On the other hand, random forest, SVM, logistic regression, Naïve Bayes, KNN, and decision tree machine learning models are trained on the depression dataset.

Once the models are trained on the datasets, model evaluation is initiated which suggests the performance of each model will be assessed using appropriate evaluation metrics. Then, hyperparameter turning is performed over two models such as random forest and gradient boosting for the cardio dataset. On the other hand, hyperparameter tuning is performed on random forests and KNN on the depression dataset. This will help to enhance or optimize the performance of models by identifying the best combination of hyperparameter values. Then, an ensembling process is performed to combine the predictions from the previously trained models, and potentially enhancing the power of prediction of the models. The evaluation process will represent all final results of the ensemble model's performance.

5 Implementation

Below is the complete implementation followed for the Application:

5.1 Cardiovascular Prediction

In the implementation phase of cardiovascular disease prediction, a systematic and cyclical procedure that involved data cleansing and exploratory data analysis (EDA) and the creation and validation of machine learning models was employed. After loading the cardiovascular dataset from Kaggle, there was a thorough data cleaning process that included removing irrelevant columns (like "id") and using the IQR-based method for handling outliers.

Following cleaning, EDA was done on the 62,505 dataset items. This brought out important knowledge concerning feature relations, the category of some variables, and numerical trait distributions.

Then the study of prediction of cardiovascular disease started with a rigorous scrubbing of the data using the IQR method for handling outliers.

Upon removals, Box Plot showed that no outliers remained. Bar plots showed that the median for $ap_hi)was130, for(ap_lo)/itwas85, forHeight, itwas165, andforWeight, itwas72.$ Further EDA also revealed significant correlations, including .52 between height and gender and .71 between $ap_hi)and(ap_lo)/.$

Categorical variables like gender, cholesterol, glucose level, cigarette smoking, alcohol drinking, or physical activity were more meaningfully viewed in terms of histogram plots.

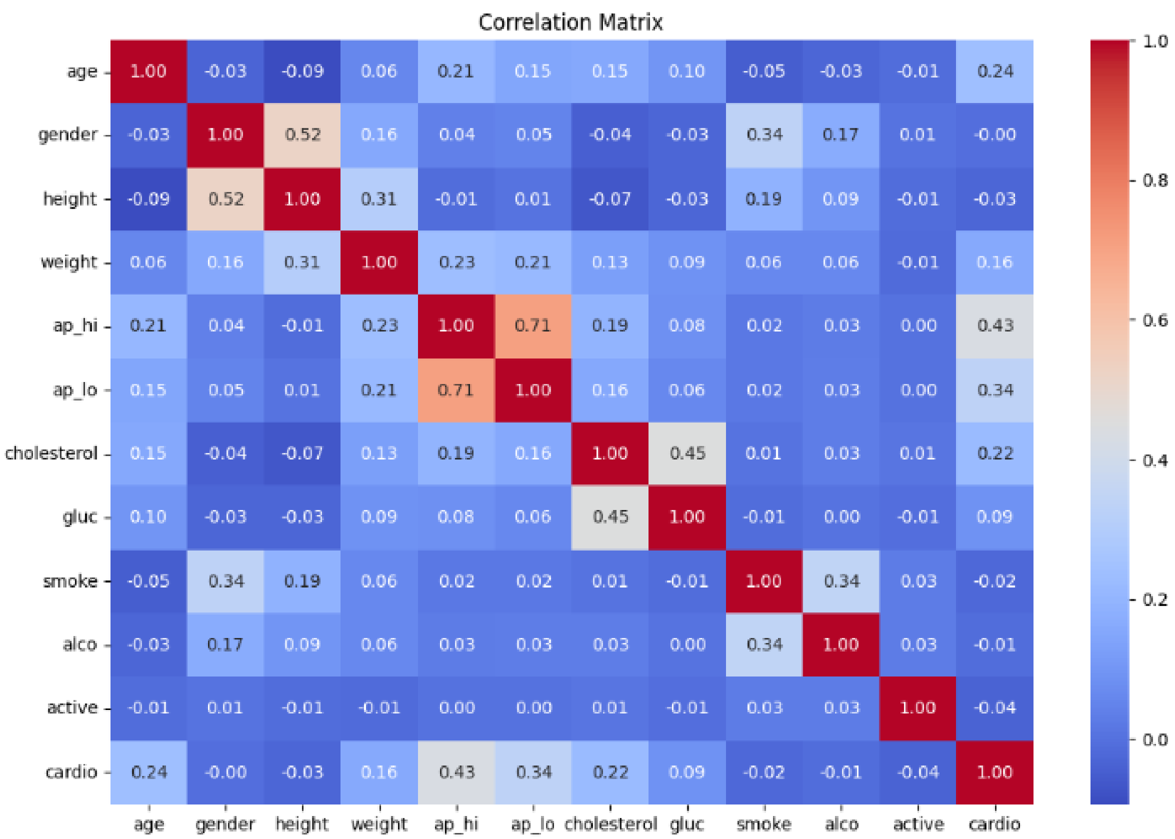

Figure 5: Corrlation Matrix

Subsequently, the data separation process saw the division of the data into training and test sets, following its segmentation into features (X) and the target variable (y). Numerical features were standardized using the StandardScaler. The model training phase utilized several classifiers, including Random Forest, Logistic Regression, Support Vector Machine, Gradient Boosting, k-nearest Neighbours, Decision Tree, and a Voting Classifier. Every output from the classifier is well examined by using confusion matrices and classification reports.

The hyperparameter tuning on Random Forest and Gradient Boosting models was aimed at improving the prediction accuracy. This entailed setting parameters like the number of estimators, max depth, and learning rate. Then, the optimal hyperparameters were used to instantiate models, and an Ensemble Model was built with a Voting Classifier. The final ensemble model was more accurate, and precise, had a higher recall, and a larger F1 score than the individual classifiers. Throughout the whole implementation, all of it was in Python, including, among others, packages like sci-kit-learn, pandas, and seaborn. The integrative approach used machine learning techniques, statistical analysis, and visualizations to predict a reliable model for cardiovascular disease. For Cardiovascular Predictions, the last output includes a prediction based on a machine-learning model for Disease or Not.

5.2 Depression Prediction

In the final stage of deployment, we have successfully created and refined machine learning models that predict depression based on socio-economic characteristics. We loaded the dataset, cleaned it up to deal with missing values, and then did EDA to see how important measures were distributed and how they correlated with depression.

After EDA we preprocessed the data by splitting the data into features and the target variable, scaling the numerical features using a standard scaler, and splitting the dataset into the training set and the testing set.

It comprised of twenty-three attributes and one thousand four hundred twenty-nine cases mostly regarding social economic and demographic parameters.

In the EDA, the use of histogram plots enabled the determination of the distributions of numerical parameters and also, the frequencies of specific risk variables. However, the association of these depression occurrences with categories characteristics like gender, marital status, level of education, and labor activities were considered.

The EDA detected some significant patterns in the dataset, including that men were more frequent, which could mean a correlation between depression and sex. Nevertheless, there was no evidence to suggest that marriage or education level were associated with depression rates. The study also considered categorical factors such as labor participation, which gave rise to an unusually high frequency in the non-labor group that raises the possibility of association with depression. The above findings offer the basis for future model description and training.

We did classifiers like, KNN, Random forest, SVM, Decision tree, Naive Bayes, and Logistic Regression. The results for these evaluation parameters include recall, accuracy, precision, and F1 score for each model. GridSearchCV provided the hyperparameters that improved the performance of the prediction of the Random Forest and KNN model.

Additionally, we employed a Voting Classifier to build an ensemble model that leveraged the strengths of random forest and KNN models. The ensemble model exhibited greater accuracy as far as precision, recall, and F1 score were concerned.

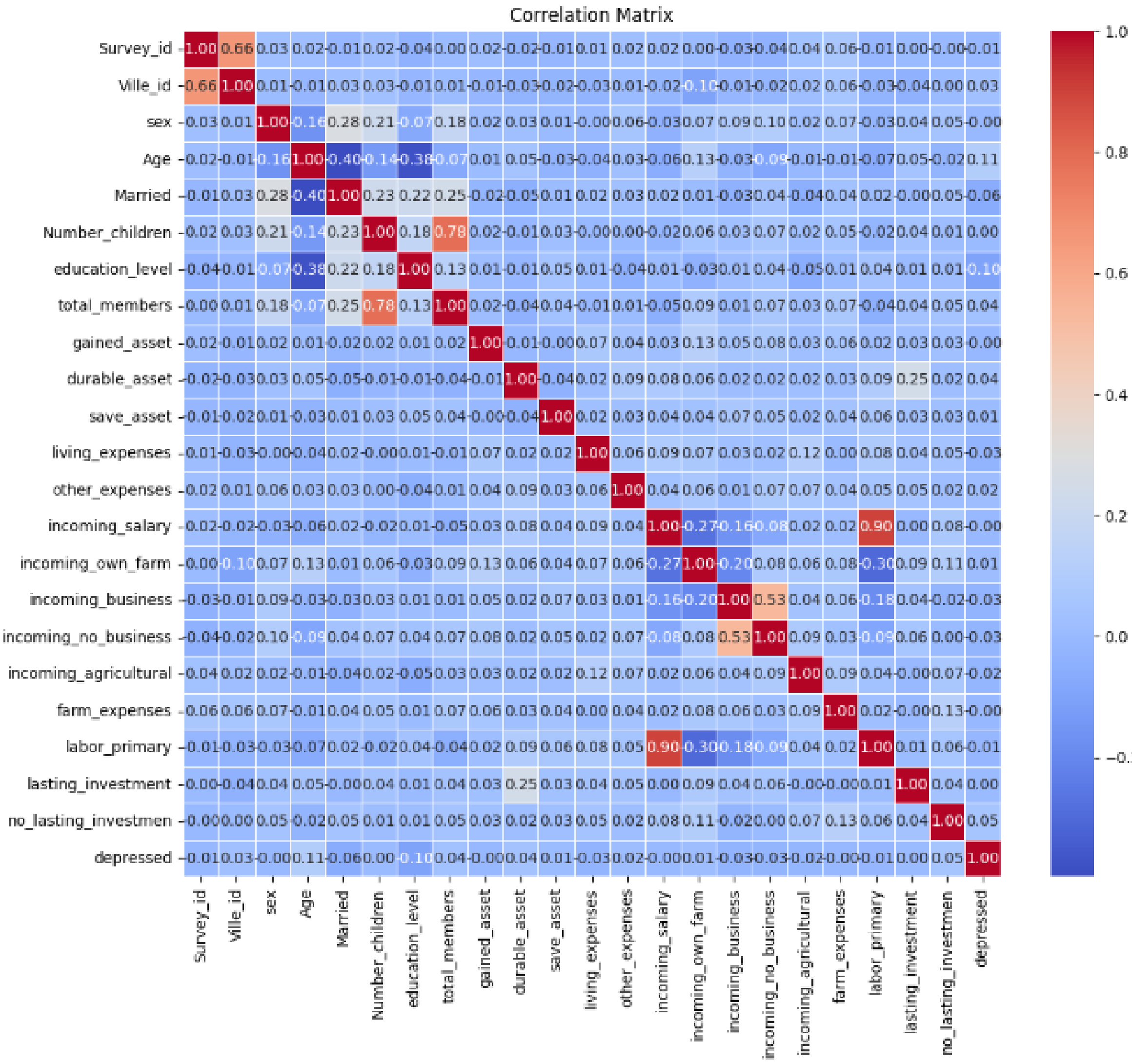

Figure 6: Corrlation Matrix

This implementation entails a detailed exploration of the model's performance, an understanding of key socioeconomic variables driving depression, and optimized machine learning models for depression prediction. For this implementation, I used Python and libraries such as sci-kit-learn, pandas, matplotlib, seaborn, and imbalanced-learn. Therefore, the final output of the Depression Prediction is the user's prediction on whether a user is depressed or not.

5.3 Flask Application

To have the suggested health prediction system implemented, a web application that uses the Flask Python framework must be developed. With the help of machine learning, it can predict cardiovascular health and depression based on user-provided data. The backend uses SQLAlchemy which provides an interface to a SQLite database for handling the login and user registration feature.

Models.py defines the data models that define the database interaction (using SQLAlchemy), and also the structure of a user database. There are several sections on the website. These include depression, prediction of cardiovascular health risks, login, and registration.

In terms of machine learning, pickle is used to load pre-trained ensemble models and scalers for cardiovascular and depression prediction. The models, which are based on the scikit-learn framework, were developed using the ensemble approach, especially for classification.

For the front-end implementation, HTML templates are applied, while for the display of dynamic content, Jinja templating is used. Through the use of CSS and the responsive nature of the user interface's design, a user-friendly experience is assured.

User authentication is done using Flask sessions, hence, users will have to register and log in to enjoy predictive capabilities. It also enhances user experience and security. It simplifies the deployment process using a lightweight development server that comes with Flask's built-in server.

The final output from a Flask Application is that a user can predict cardiovascular disease and depression. The last stage in the process will include users who may forecast various health-related difficulties with ease using a smooth and user-friendly platform through web forms and the system will make predictions based on the developed machine learning models. These three technologies, namely, web development, machine learning, and backend database management come together to yield a complete and useful health prediction system.

6 Evaluation

Cardiovascular and Depression Machine Learning models were evaluated as below:

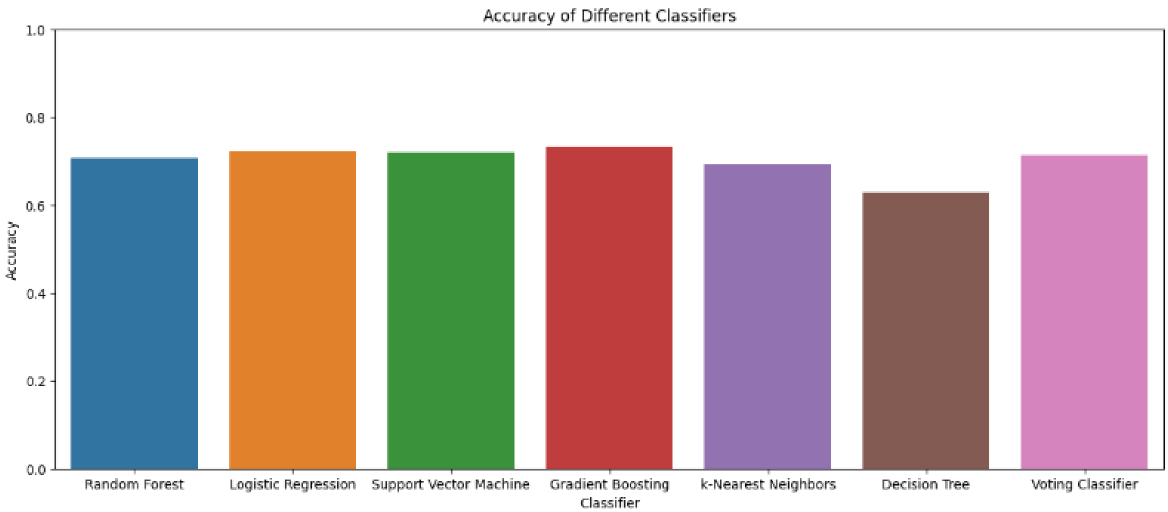

Figure 7: Accuracy of Different Classifiers in Cardiovascular Disease Prediction

6.1 Cardiovascular Prediction

After data preparation, training and test sets were extracted from the dataset and then standardized. We developed and carefully assessed seven classifiers using accuracy, precision, recall, and F1 score metrics:

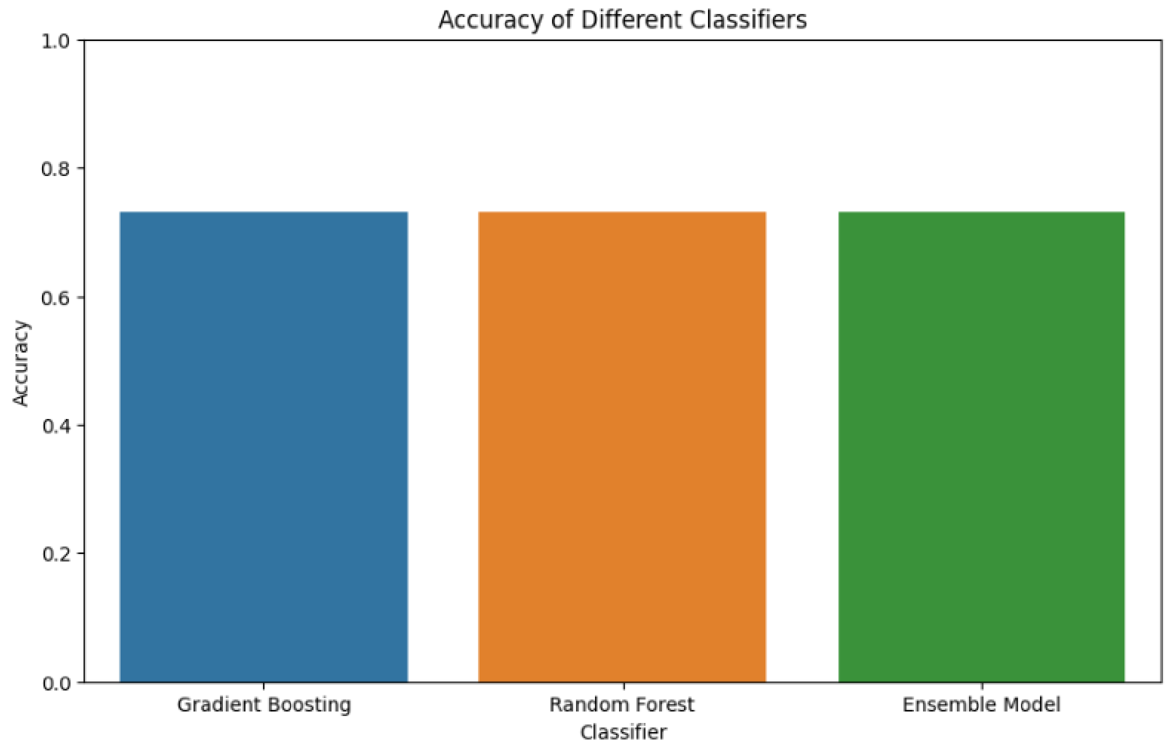

Figure 8: Accuracy of Hypertuned Models in Cardiovascular Data Prediction

A voting classifier ensemble, decision tree, random forest, gradient boosting, SVM, logistic regression, and KNN. However, Logistic regression ranked as the best performer with a 72 Percent accuracy that surpassed most of the individual classifiers. SVM had an accuracy of 72 Percent and balanced precision and recall.

Model	Accuracy
Random Forest	0.71
Logistic Regression	0.72
SVM	0.72
Gradient Boosting	0.73
KNN	0.69
Decision Tree	0.63
Voting Classifier	0.71
Ensembled voting (Random forest gradient boosting hypertuned)	0.73

Figure 9: Accuracy results of Cardio

Through fine-tuning of the Random Forest and Gradient Boosting algorithms, optimized configurations were achieved for the advancement of the application in question. A voting classifier was used to create an ensemble model that consisted of a random forest model, an adjusted gradient boosting model, and a random forest model that had been tweaked. This ensemble model yielded a respectable 73 Percent accuracy. The ensemble model showed good predictive power for both cardiovascular disease groups, with a significant increase in recall and accuracy. The evaluation of the performance of these classifiers revealed that Random Forest was 71 percent, Logistic Regression was 65 percent, SVM was 68 percent, Gradient Boosting was 76 percent, KNN was 49 percent and Decision Tree was 54 percent. The voting classifier's ensemble achieved 71 percent accuracy. These discoveries suggest that ensemble methods and model selection are crucial for predictive modeling in predicting cardiovascular disease. These findings underscore the importance of a balanced strategy for optimizing recall and accuracy metrics yielding robust predictions.

6.2 Depression Prediction

Model	Accuracy
Random Forest	0.82
Logistic Regression	0.83
SVM	0.78
Gradient Boosting	0.83
KNN	0.75
Decision Tree	0.81
Voting Classifier	0.82
Hyper-tuned KNN	0.82
Hyper-tuned Random forest	0.83
Ensemble	0.83

Figure 10: Accuracy results of Depression Results

Classification Report for Ensemble (KNN + Random Forest):					
		Precision	Recall	F1 score	Support
	0	0.84	0.96	0.90	239
	1	0.31	0.09	0.13	47
accuracy				0.82	286
macro avg		0.58	0.52	0.52	286
weighted avg		0.75	0.82	0.77	286

Figure 11: Classification Report For Ensemble(KNN+Random Forest) on Depression dataset

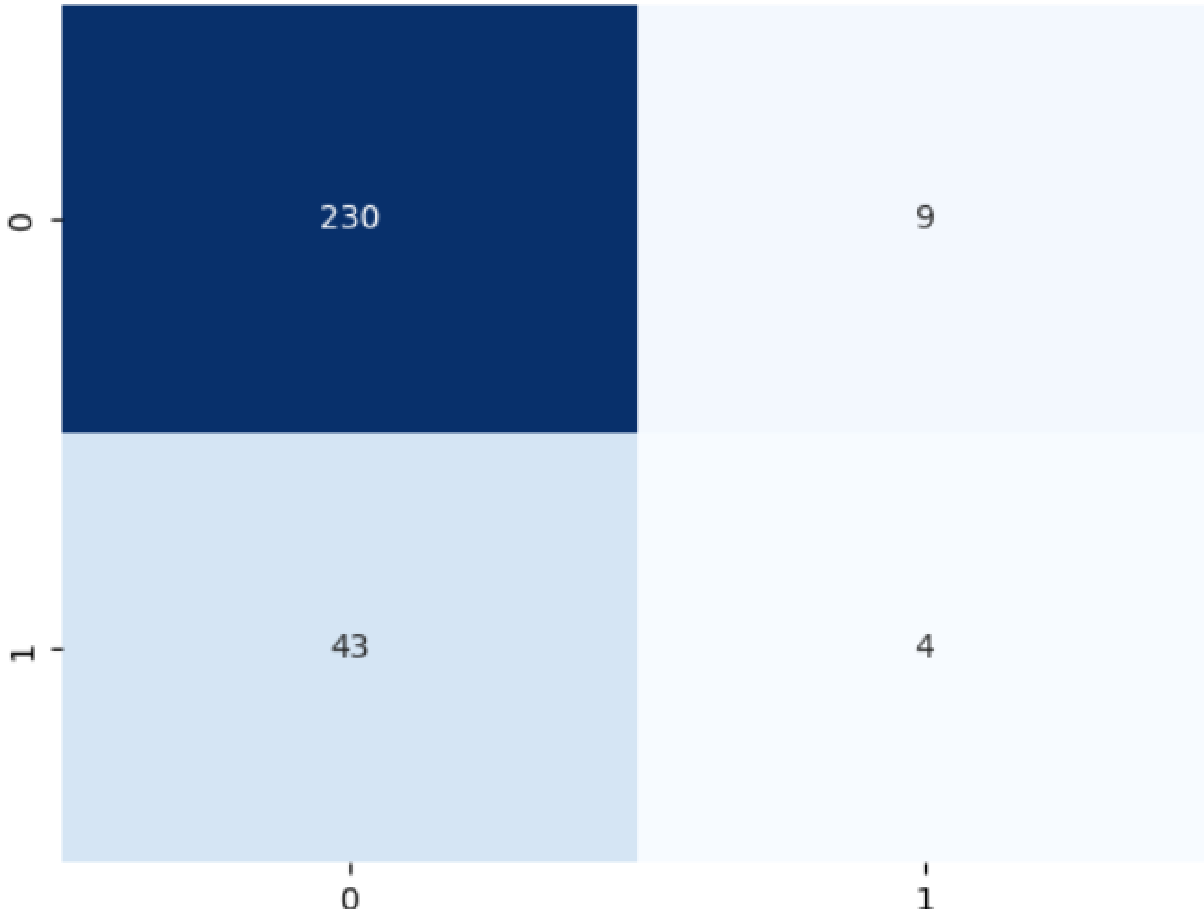

Figure 12: Confusion matrix

The depression prediction research commenced with an intensive survey into the dataset which entailed data cleaning, missing value replacement, and an in-depth analysis of significant statistical factors.

The EDA revealed significant trends on the dataset including the higher frequencies of men suggesting a probable relationship between depression and gender. The results showed that marital status and level of education did not influence the prevalence of depression. This also included categorical features such as "work involvement", where a high frequency in the non-labor group could potentially signify a correlation with depression. Thus, these results are needed for the subsequent model interpretation and training.

The depression prediction task was undertaken using a dataset with socioeconomic and demographic features that were fed into multiple machine-learning techniques. To predict depression, seven classifiers were used: Algorithms like Random Forest, Logistic Regression, Naive Bayes, SVM, Decision Tree, and KNN. The Random Forest model scored 7.41 percent F1, 4.26 percent recall, 28.57 percent precision, and 82.52 percent accuracy. The Logistic Regression was accurate, with the F1 score of 83.22 percent. Although it presented a provisionality = 0 percentage, indicating possible limitations. Naive Bayes achieved a precision of 22.22 percent, a recall of 12.77 percent, and an F1 score of 16.22 percent, resulting in an accuracy of 78.32 percent.

SVM obtained an F1 score of 0.00 percent, with accuracy, precision, and recall at 83.57 percent. Its accuracy was moderate at 75.17 percent and the precision, recall, and F1 score for the Decision Tree model were 23.91 percent, 23.40 percent, and 23.66 percent, respectively. The prediction accuracy for KNN was 81.12 percent, recall = 4.26 percent, precision = 18.18 percent, and F1 score = 6.90 percent. With hyperparameter tuning, KNN's accuracy improved to 82.17 percent, recall of 2.13 percent, precision of 16.67 percent, and F1 score of 3.76 percent.

Apart, the optimal accuracy in hyperparameter optimization of the Random Forest model was 83.22 percent, recall of 4.26 percent precision of 40.00 percent, and F1 score of 7.84 percent. For example, the ensemble method used Random Forest and KNN with Voting Classifier to achieve an F1 score of 4.00 percent, accuracy of 83.22 percent, precision of 33.33 percent, and recall value of 2.13. In summary, these results portray the tradeoffs between the precision, recall, and overall accuracy of the depression prediction and provide an overview of each model's predictive performance.

6.3 Flask Application

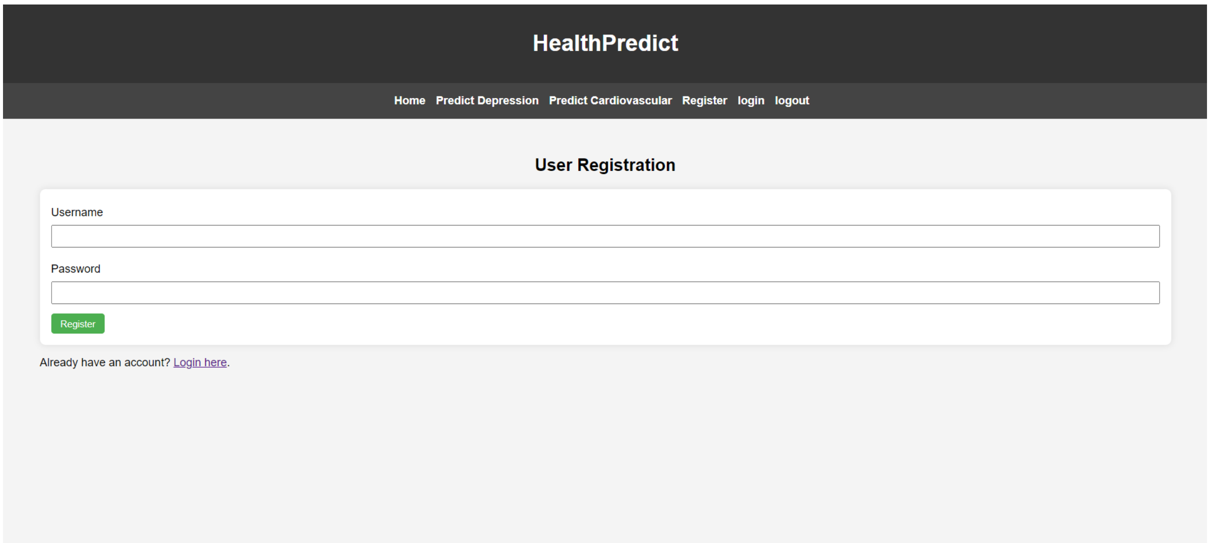

Figure 13: User Registration Page

The combination of two machine learning models, one for the prediction of depression and the other, for that of Cardiovascular, was applied in the assessment of the Flask application. The program is designed intuitively with registration, login, and logout features to provide a personalized experience. Security measures are put in place to prevent any unauthorized login or registration.

The cardiovascular prediction module allows users to enter multiple health factors which are fed into a pre-fitted scalar tailored for the cardiovascular characteristics. The user interface shows the predicted response of the ensemble model for cardiovascular health. Moreover, the program handles the user sessions in such a way that every logged-in individual can only view the predictions.

Similarly, the depression prediction module captures socio-demographic information which is scaled through a pre-fitted scaler for depression-related characteristics. The user is provided with the results of using the ensemble model for depression prediction. For the registration and login methods, they access a SQLite database to safely manage the user's information.

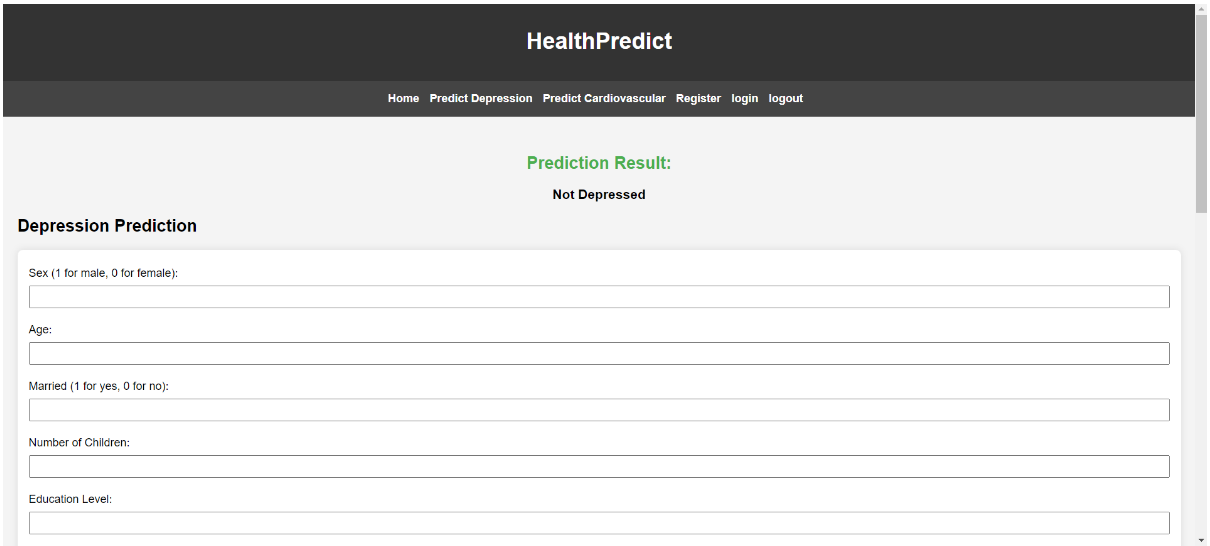

Figure 14: Prediction Result for Depression for a Sample Input

Error management for duplicate usernames during registration and unsuccessful login attempts makes up a part of this system. In addition, the site boasts of a simple homepage, which displays a choice of either cardiovascular or depression predictors with corresponding input forms.

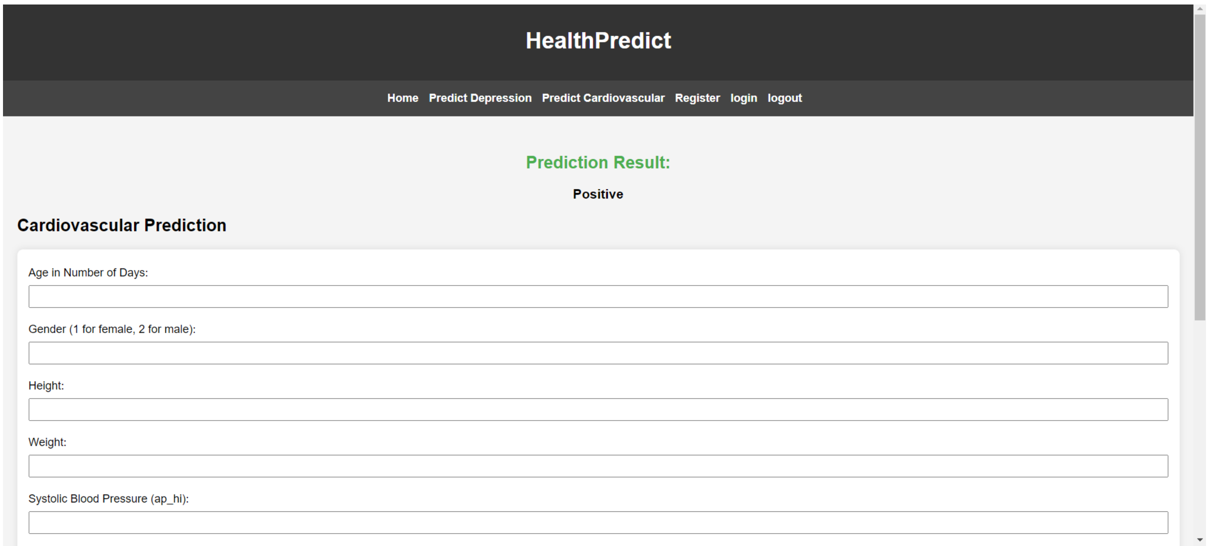

Figure 15: Prediction Result for Cardiovascular for a Sample Input

Customers can easily use this Flask application which has incorporated machine learning algorithms for predicting mental and cardiovascular health. At the end of the day, the technology delivers a seamless user experience via a well-thought-out interface that offers practical healthcare predictions based on the power of machine learning.

6.4 Discussion

The evaluation of models for predicting cardiovascular illness provided important insights into the performance of various classifiers. Logistic regression turned out to be a highly efficient method with a 72 percent accuracy rate. The above result proves the strength of logistic regression in prognosis of the cardiovascular diseases and agrees with the other studies on using this technique in medical care. The SVM model had a balanced precision and recall, implying that it could predict across illness and non-disease categories. Interestingly, it also recorded an impressive accuracy of 72 Percentage.

It was shown that ensemble approaches are capable of increasing the predicting accuracy, specifically by applying Random Forest and modified Gradient Boosting models in a Voting Classifier. As shown by the result, the ensemble model achieved an astonishing 73 Percent accuracy rate compared to single classifiers. It strengthens the importance of ensemble techniques in achieving precision in heart disease prediction. A performance study revealed each classifier's strong and weak points, with Gradient Boosting being the most accurate at 73 Percent.

The Random Forest, however, was a good model for predicting depression in terms of accuracy with 83.22 Percentage F1 score of 7.41 Percentage precision of 28.57 Percentage, and recall with 4.26 Percentage. Logistic Regression had a lower precision and recall of 0.00 Percentage whilst showing an 83.22 Percentage accuracy. This reflects the fact that this study was not fair. It was evident that precision-recall-and-total accuracy trade-offs provided information regarding the predictive power of each model. Instead, the hypertuned model was chosen as opposed to Smote, which gave a poor F1 score. This could be due to inconsistent data.

Additionally, the use of Hyperparameter Optimization led to an improvement in some models like the Random Forest model which shows that fine–tuning the parameters is necessary for better results. The ensemble approach (Random Forest + kNN through Voting Classifier) had an F1 score of 4.00 Percentage accuracy of 83.22 Percentage. This highlighted the challenge of optimizing multiple models for depression prediction.

The inclusion of depression and cardiovascular prediction algorithms into a user-friendly system made the Flask application successful. Scalability evaluation, secure data processing, and user interface design also affect a good user experience. This program provides simple-to-use utilities that enable the user to track his cardio and mental health, according to the current trend regarding personalized healthcare forecasting using machine learning.

Our results provide further support to earlier studies within the literature review that point to the growing importance of ensemble approaches and personalized health forecasting models. Machine learning models must be made adaptive to changes in healthcare data and developments in predictive analytics. Further improvements might involve looking for more features, incorporating more elaborate algorithms, and considering possible biases in the data. In summary, these results underscore the importance of continuous evaluation and improvement in predictive modeling and contribute to the ongoing discourse on the employment of machine learning in healthcare.

7 Conclusions and Future Work

This study had objectives of evaluating prediction models, developing an app based on FLASK, and criticizing the results to address the research problem of employing machine learning for depression/cardiovascular prediction in healthcare. The study's objectives were adequately met, providing information on the accuracy of machine learning prediction about depression and cardiovascular diseases.

The main conclusions of this evaluation pointed out the strength of logistic regression in predicting cardiovascular outcomes that gave an accuracy of 72 Percent. In addition, the Random Forest and modified Gradient Boosting ensemble model exhibited an impressive prediction accuracy of 73 Percent. The Random Forest model outperformed the other models by predicting depression with an accuracy of 83.22 Percentage. Thus, it proved to be effective in the domain of mental health. This demonstrates how machine learning can accurately predict the challenges of heart and mental issues.

Important implications of the research can thus be translated into a usable application via the creation of an intuitive Flask application. Machine learning models introduced into a platform that assesses mental health and heart diseases can enhance personalized healthcare. Therefore, one must also be aware of the restrictions, e.g., a necessary large amount of data as well as existing limitations between recall, precision, and general level of accuracy.

However, future research should focus on improving upon the prediction models. However, tackling some of the potential biases in the data set, developing more advanced algorithms as well as researching more relevant parameters could bring us closer to these meaningful gains. Additionally, owing to the evolving field of predictive analytics, future investigations may consider the dynamic nature of healthcare data. The models can be made to complement clinical processes by involving medical experts in future work plans and extending their coverage to other health factors.

Furthermore, given the increasingly urgent requirements for tailor-made health applications, the opportunity for commercialization should also be looked into. In this regard, collaborations with technology firms and healthcare providers may facilitate the integration of the predictive models in current health systems. A validation procedure and respect for ethical practices, however, should support commercialization initiatives to ensure accurate prediction models' dependability and credibility in authentic healthcare settings. In essence, this study provides a basis for future efforts toward making machine learning models more accurate and appropriate in the detection and management of medical conditions.

References

Alty, S., Millasseau, S., Chowienczyc, P. and Jakobsson, A. (2004). Cardiovascular disease prediction using support vector machines, Vol. 1, pp. 376 – 379 Vol. 1.

Amma, N. G. B. (2012). Cardiovascular disease prediction system using genetic algorithm and neural network, *2012 International Conference on Computing, Communication and Applications*, pp. 1–5.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., Cannon, T. D., Krystal, J. H. and Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach, *The Lancet Psychiatry* **3**(3): 243–250.
URL: *https://www.sciencedirect.com/science/article/pii/S221503661500471X*

Collaboration, E., E, D., Gao, P., Pennells, L., S, K., Caslake, M., A, T., AS, B., N, S., Wormser, D., Saleheen, D., CM, B., BM, P., Sundström, J., PM, R., D, N., RF, G., I, F. and P, D. (2012). Lipid-related markers and cardiovascular disease prediction, *JAMA The Journal of the American Medical Association* **307**: 2499–506.

C-Reactive Protein, Fibrinogen, and Cardiovascular Disease Prediction (2012). *New England Journal of Medicine* **367**(14): 1310–1320. PMID: 23034020.
URL: *https://doi.org/10.1056/NEJMoa1107477*

Dinesh, K. G., Arumugaraj, K., Santhosh, K. D. and Mareeswari, V. (2018). Prediction of cardiovascular disease using machine learning algorithms, *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1–7.

Elayan, H., Aloqaily, M. and Guizani, M. (2022). Sustainability of healthcare data analysis iot-based systems using deep federated learning, *IEEE Internet of Things Journal* **9**(10): 7338–7346.

Firdaus, H., Hassan, S. and Kaur, H. (2018). A comparative survey of machine learning and meta-heuristic optimization algorithms for sustainable and smart healthcare.

Gao, S., Calhoun, V. D. and Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction, *CNS neuroscience amp; therapeutics* **24**(11): 1037—1052.
URL: *https://europepmc.org/articles/PMC6324186*

Hawes, M. T., Schwartz, H. A., Son, Y. and Klein, D. N. (2023). Predicting adolescent depression and anxiety from multi-wave longitudinal data using machine learning, *Psychological Medicine* **53**(13): 6205–6211.

Jain, K. and Kaushal, S. (2018). A comparative study of machine learning and deep learning techniques for sentiment analysis, *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 483–487.

Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H., Kaplin, S., Narasimhan, B., Kitai, T., Baber, U., Halperin, J. L. and Tang, W. H. W. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis, *Scientific Reports* **10**: 16057.

Lamiae, E., Hicham, G., Fatiha, E. and Mohammed, B. (2022). Smart home and machine learning as a sustainable healthcare solution: Review and perspectives, *The Proceedings of the International Conference on Smart City Applications*, Springer International Publishing, Cham, pp. 145–155.

Mcmahon, M., Hahn, B. and Skaggs, B. (2011). Systemic lupus erythematosus and cardiovascular disease: Prediction and potential for therapeutic intervention, *Expert review of clinical immunology* **7**: 227–41.

Patel, M., Khalaf, A. and Aizenstein, H. (2015). Studying depression using imaging and machine learning methods, *NeuroImage: Clinical* **10**.

Priya, A., Garg, S. and Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms, *Procedia Computer Science* **167**: 1258–1267. International Conference on Computational Intelligence and Data Science.
URL: *https://www.sciencedirect.com/science/article/pii/S1877050920309091*

Sajjadian, M., Lam, R. W., Milev, R., Rotzinger, S., Frey, B. N., Soares, C. N., Parikh, S. V., Foster, J. A., Turecki, G., Müller, D. J., Strother, S. C., Farzan, F., Kennedy, S. H. and Uher, R. (2021). Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis, *Psychological medicine* **51**(16): 2742—2751.

Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T. and Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?, *Heart* **104**(14): 1156–1164.
URL: *https://heart.bmj.com/content/104/14/1156*

Usman, M., haris, s. and Fong, A. (2020). Prediction of depression using machine learning techniques: A review of existing literature.