**School of Computer Science and Mathematics**

A logo with a book and text

Description automatically generated

**Student ID: 21025647**

**Module Name:** Applications of Artificial Intelligence, Machine Learning and Data Science

**Module Number:** CSC-40070

**Lecturer:** Dr Shaily Kabir

**Group Members**: Millicent Djabatey (17018405) and Angel Ariyibi (21025647)

**Word Count: 3830** (excluding, figure legends, headings, appendix, references, and footnotes)

## Introduction

Machine learning is a rapidly growing field in artificial intelligence that enables computers to learn from data and make predictions or decisions [4] [6]. There are several distinct techniques in machine learning, primarily categorized into four main types: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [4].

In Supervised Learning, the algorithm is trained on labelled data, where each input is paired with the corresponding correct output. This type of learning aims to learn a mapping function from input variables to output variables. Conversely, Unsupervised Learning involves training the algorithm on unlabelled data, where the algorithm tries to learn the underlying structure or distribution in the data. Semi-supervised learning combines elements of both supervised and unsupervised learning, where the algorithm is trained on a combination of labelled and unlabelled data, often leveraging the abundance of unlabelled data to improve learning accuracy. Lastly, Reinforcement Learning involves training an agent to make sequential decisions by interacting with an environment, where the agent receives feedback in the form of rewards or penalties based on its actions. While all these types of learning are important in machine learning, this paper will focus specifically on one type of supervised learning known as classification [2] [4] [3][6].

Classification is a type of supervised machine learning task where the goal is to predict the class or label of new data points based on learned patterns from a training dataset containing labelled examples [6]. Some of the main applications of classification span across multiple domains:

* Healthcare: Predicting diseases based on patient symptoms, clinical data, and relevant features. For example, classifying whether a patient has cancer or not based on medical test results.
* Email Services: Classifying emails as spam or not spam (binary classification).
* Cybersecurity: Detecting and classifying various types of network attacks like Denial of Service, User to Root attacks, etc. (multi-class classification).
* Text Classification: Categorizing text documents, news articles, etc. into multiple hierarchical classes like city, technology, latest news (multi-label classification).
* Finance (loan approval prediction): Analysing historical loan data to predict whether new loan applications should be approved or denied based on factors like income, credit history, loan amount, and purpose, enabling institutions to streamline approval processes and minimize risk while enhancing customer satisfaction.

Classification problems can be broadly categorized into binary (two classes), multi-class (more than two classes), and multi-label (assigning multiple non-exclusive labels) tasks. The goal is to learn a mapping function from input features to the output class labels based on patterns in the training data [4] [5]. Classification has wide-ranging applications across industries for predictive modelling and decision-making. This paper is focused on the application of classification in healthcare - disease diagnosis.

The task of predicting diseases based on patient symptoms, clinical data, and other relevant features is a common problem in the healthcare industry. Disease prediction algorithms play a crucial role in early detection, prognosis, and treatment planning, ultimately improving patient outcomes and reducing healthcare costs [7]. By analysing large volumes of patient data, including medical history, genetic information, laboratory test results, and imaging studies, machine learning algorithms can identify patterns and relationships that may not be immediately apparent to human clinicians. This enables healthcare providers to make more accurate diagnoses, predict disease progression, and personalize treatment plans based on individual patient characteristics. Moreover, disease prediction algorithms can help prioritize healthcare resources, identify high-risk populations for preventive interventions, and guide public health initiatives aimed at disease prevention and control [9].

Again, in situations where access to healthcare resources is limited or where rare diseases are involved, analysing the patterns in the patient's symptoms using disease prediction algorithms can suggest potential diagnoses, which can then be further evaluated and confirmed by medical experts [8] [11].

## 2. Challenges in Disease Classification Algorithms.

Developing an effective machine learning algorithm for disease prediction and in healthcare in general involves several significant challenges that need to be addressed. Some of the key challenges are:

* **Data Quality:** Data quality is paramount in developing effective machine learning algorithms for disease prediction. Medical data, often noisy, incomplete, or inconsistent, poses a significant challenge to model training. Missing values, inaccurate measurements, and inconsistent coding practices can introduce errors and biases into the dataset, undermining the model's performance [16].
* **Imbalanced Datasets:** Dealing with imbalanced datasets presents a significant challenge in developing machine learning algorithms for disease prediction within the healthcare domain. Often, medical datasets exhibit highly skewed distributions, with certain diseases being more prevalent than others. This imbalance can result in models that are biased towards the majority class, leading to poor performance in predicting minority class instances, such as rare diseases[10].
* **Feature Selection:** Medical datasets are typically rich in features, including symptoms, test results, and demographic information, among others. However, not all features may contribute equally to the predictive performance of the model, and some may even be irrelevant or redundant. Feature selection poses a critical challenge in developing machine learning algorithms for disease prediction in healthcare [14].
* **Model Interpretability:** In the healthcare domain, ensuring the interpretability and explainability of machine learning models is crucial. Medical professionals require insight into the reasoning behind the model's predictions to make informed decisions. However, many popular machine learning algorithms, such as deep neural networks or ensemble methods, are often perceived as "black boxes," lacking transparency in their decision-making process. Providing interpretable outputs is essential in empowering healthcare professionals to trust and effectively utilize machine learning predictions in clinical practice to help improve patient care and outcomes [11]. Therefore, exploring and developing techniques such as feature importance analysis and partial dependence plots is of great importance in this field.
* **Handling Rare Diseases:** Addressing the challenge of handling rare diseases presents a significant hurdle in the development of machine learning algorithms for disease prediction in healthcare. Certain diseases are inherently rare, and the available data for these conditions may be limited. When training machine learning models on imbalanced datasets, the scarcity of data for rare diseases can lead to biased models that struggle to accurately predict these cases, potentially resulting in misdiagnosis or delayed treatment for affected patients [12].
* **Ethical Considerations:** Ethical considerations loom large in the development and deployment of disease prediction models within healthcare. These models carry profound implications, as inaccuracies or biases within them can result in detrimental outcomes for patients. Concerns surrounding data privacy, consent, and the potential for biases—such as those related to race, gender, or socioeconomic status—demand meticulous attention throughout the entire process of model development and deployment. Regulatory compliance and adherence to ethical guidelines are imperative to guarantee the responsible and equitable use of disease prediction models in healthcare settings [13].

To overcome these challenges, a multidisciplinary approach, involving collaboration between machine learning experts, domain experts (medical professionals), and ethical oversight is necessary to develop reliable, interpretable, and fair disease prediction models that can be effectively integrated into clinical decision-making processes. Again, the implementation of pre-processing techniques, feature importance analysis, oversampling and undersampling and many more can help mitigate these issues.

## 3. Approaches in Disease Prediction/ Classification

Different classification algorithms and techniques can be used in the development of algorithms for disease classification. Among the most widely used approaches are machine learning algorithms (including logistic regression, naive Bayes, decision trees, random forests, support vector machines), deep learning architectures(DLAs) like neural networks and hybrid methods of the various approaches. Each approach has its own strengths, weaknesses, and underlying assumptions, making them suitable for different scenarios and problem domains. This section aims to compare some of these approaches.

* **Machine Learning Algorithms**:

Several machine learning algorithms have been employed in the development of disease prediction algorithms to aid clinical decisions. Logistic regression is a statistical model that uses a logistic or sigmoid function to estimate the probabilities of class membership. It works well when the data is linearly separable but can overfit high-dimensional datasets, necessitating regularization techniques like L1 and L2. Naive Bayes, on the other hand, is based on Bayes' theorem with the assumption of independence between features. It requires a small amount of training data and is computationally efficient, making it suitable for various real-world applications like text classification and spam filtering. However, its performance can be affected by the strong assumption of feature independence [3][9].

Decision trees and random forests are non-parametric supervised learning methods that can handle both classification and regression tasks. Decision trees recursively split the data based on feature values, creating a tree-like model that is easy to interpret but prone to overfitting, especially with deep trees or noisy data [3] [8]. Random forests mitigate this issue by combining multiple decision trees, using techniques like bagging and random feature selection, resulting in improved performance and robustness to noise and outliers[3].

Support vector machines (SVMs) take a different approach by constructing hyperplanes or sets of hyperplanes to separate classes in high-dimensional space. They aim to find the maximum margin hyperplane that best separates the classes, making them effective in high-dimensional spaces and capable of using different kernel functions (linear, polynomial, RBF, sigmoid, etc.). However, SVMs may struggle with overlapping classes or noisy data [14].

* **Deep Learning Architectures (DLAs):**

DLAs, particularly neural networks, have gained significant attention in recent years due to their ability to learn complex patterns from high-dimensional and non-linear data. By leveraging convolutional neural networks and other deep learning models, researchers can analyse medical images, genomic data, and clinical records to accurately diagnose various diseases. These architectures excel at identifying subtle features and correlations that may be overlooked by traditional methods. However, challenges such as data quality, model interpretability, and ethical considerations must be addressed to ensure responsible deployment of these powerful techniques in healthcare settings[17][20].

* **Hybrid Approach**

Another approach that can be used in disease prediction algorithms is the combination of multiple technique or models to leverage their strengths. Hybrid approaches in disease prediction algorithms combine the strengths of multiple machine learning techniques to enhance predictive accuracy and robustness. For instance, decision trees are great at feature selection and interpretable rule generation [3] [8], while deep neural networks excel in capturing complex patterns from high-dimensional data[17]. Again, ensemble methods, such as stacking or boosting, often yield superior performance compared to standalone methods by integrating predictions from individual models used in their development [18]. Hybrid approaches provide flexibility in adapting to different dataset characteristics and disease domains, making them valuable tools in medical decision-making and patient care [20].

## 4. Selected Approach- Stacking Classifier (Decision Tree and Neural Networks)

Considering the need for a robust and adaptable approach, our team opted for an ensemble method that leverages the combination of decision trees and neural networks aligning with the current literature acknowledging their adaptability in clinical settings [14]. Our chosen approach involved using the stacking method, where the stacking classifier combined the strengths of decision trees and neural networks to make predictions.

Stacking is an ensemble machine learning method that combines the predictions of two or more base models by training them on the same data set then use their predictions to train another model known as the meta-model to help make the final prediction [18]. Thus, stacking is a two-stage process:

* First, the base models, which could be any ML algorithm, are selected and trained on the data. In this project, multiple classification models, including logistic regression, decision tree, random forest, SVM, Naïve Bayes, and artificial neural networks (ANN), were tested and evaluated. Following that, neural networks and decision trees were selected as base models.

Decision trees offer superior interpretability and adaptability in clinical settings. They segment the decision space into smaller areas, providing explanations for each prediction [3] [8]. Decision trees were selected in this project as the first base model so they could be used for feature importance analysis, which is crucial in clinical diagnosis. Despite their efficiency in classifying unknown datasets, challenges arise with increasing variable complexity and sensitivity to training data variations [9]. To compensate for these limitations, deep learning methods, such as neural networks, were selected as the second base model in this project. Neural networks are known for their ability to capture complex nonlinear relationships and handle high-dimensional data effectively, making them suitable for medical applications with intricate patterns and large datasets [10][15] [20].

* The second stage is training the meta- model, which is logistic regression in this project, on the generated predictions of the selected base models. Logistic regression is a widely used statistical technique for binary classification problems[3], making it a suitable choice for the meta-model in this case.

Stacking is advantageous in many ways including improved accuracy by leveraging the strengths of each base model and combine them to improve overall predictive performance, robustness, and flexibility. Although very helpful and better than other ensemble methods, stacking can be quite difficult to implement and may require more computational power [18]. Again, for the best results, different base models are required for training and lots of data. **Figure 1** below shows the algorithmic process of the Stacking ensemble method.

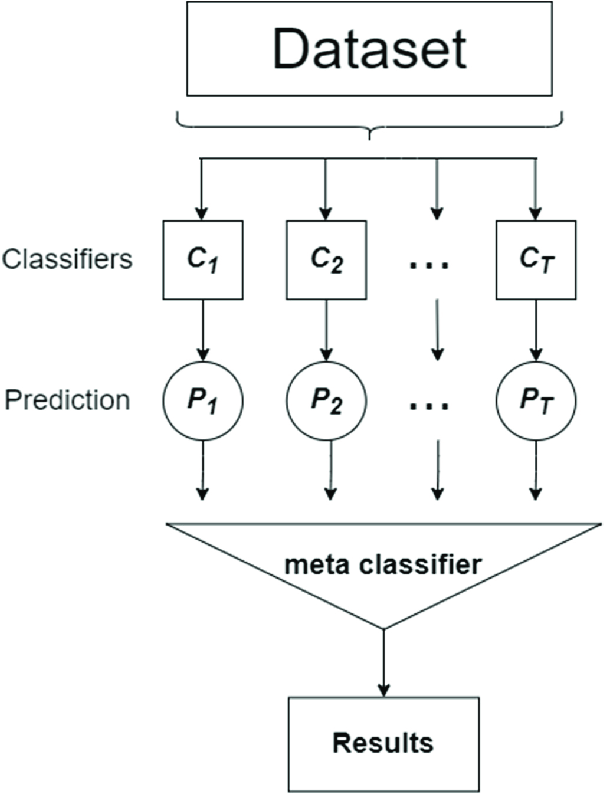


Figure 1. Algorithmic process of Stacking Ensemble Model.

## 5. Implementation of Selected Approach

**Dataset Description:**

The dataset used in this project contains two files: **Training.csv** and **Testing.csv both** of which were used for the training and testing of machine learning models respectively. The CSV file named "Training.csv" contains 4,920 data points with 133 columns in total of which 132 columns represent the different linked symptoms. The datapoints reflect the information about various symptoms and features, with the last column representing the "prognosis" or diagnosis for each data point. The dataset is balanced, with each of the 41 unique diseases having 120 rows of data.

The “Testing.csv” file on the other hand contains 42 data points, with columns like that of the Training csv. Although smaller in size, it is a set of data used to evaluate the effectiveness of the trained model.

**Pre-processing Steps:**

To ensure that the data used in the development of the model was clean, of good quality, and integrity, several data cleaning and pre-processing steps were undertaken. The key pre-processing steps are outlined below.

1. **Checking the Dataset**: The first step was to ensure that the loaded dataset was the appropriate one using Python's **head ()** method. This step is crucial because working with the wrong dataset could lead to erroneous results and invalid conclusions.
2. **Removing Unnecessary Columns**: An unnamed column was identified and removed from the dataset using **train\_data.drop('Unnamed: 133', axis=1, inplace=True**). Removing irrelevant columns is essential because using them could skew the results of the model and affect its performance.
3. **Handling Missing Values**: The code checked for any null or missing values in the dataset using the ‘**df.isnull().sum()’** function. Although the dataset used in this project had no missing values, handling missing data is a critical aspect of data analysis and machine learning. Developing models with missing values can be challenging or impossible, as the dataset cannot be trusted for its integrity and accuracy, potentially leading to stakeholders making decisions based on incorrect information.
4. **Checking for outliers**: The data was checked for outliers using box plots. The **‘temp\_df=train\_data.iloc[:,:-1]’** was used to create a new dataframe containing all the features of the main data frame (train\_data) except for the ‘prognosis’ column which was the target variable in this dataset. A box plot was then plotted to visualise the distribution of the data, to allow for the identification of outliers. Outliers are extreme values or variables found in a dataset that have the potential to negatively affect results of an analysis. There were no outliers in this dataset, nonetheless, it is important to recognise their effect on a project like this.
5. **Encoding Categorical Variables**: One of the pre-processing techniques used was converting the string values in the 'prognosis' column to numerical values using OneHotEncoder:

A screenshot of a computer code

Description automatically generated

Figure 2. Code snippet showing the implementation of Encoding process.

This step is crucial because many machine learning algorithms cannot work directly with categorical data and encoding them into numerical values is necessary for the models to process the data correctly.

1. **Feature Selection**: The feature selection process involved several steps to refine the dataset and improve the model's performance and interpretability. First, highly correlated features were identified and removed through correlation analysis, as they can introduce multicollinearity issues and lead to unstable model estimates. Next, variance thresholding was applied to eliminate low-variance features that may not contribute significantly to the learning process, thereby reducing dimensionality and preventing overfitting. Additionally, Pearson correlation and feature importance analysis were conducted to gain insights into the relationships between variables and their respective contributions to the model's predictions. By analysing the correlation matrix and feature importance scores, redundant features could be identified and removed, further enhancing the model's efficiency and interpretability[21] [22].

A screenshot of a computer program

Description automatically generated

Figure 3. Code snippet showing the implementation of the different validation processes.

**c. Extracted features :**

For this medical diagnosis model, the relevant features extracted from the dataset include various symptoms and characteristics related to different diseases. The dataset contains numerous columns representing the various symptoms a patient may experience for a specific disease. These features serve as input variables for the machine learning models to learn patterns and make accurate predictions about the disease or condition being diagnosed.

Firstly, we used two feature selection techniques - correlation-based feature removal and variance thresholding, to identify and retain the most informative and non-redundant features from the initial set of variables. The selected features produced were then used to train and evaluate the performance of different machine learning algorithms, including decision trees and neural networks, the chosen approaches for this disease classification tasks.

All the models achieved perfect or near perfect accuracy, so we used different validation and evaluation techniques such as cross-validation, confusion matrix, hyper-parameter tuning- specifically - Grid search and voting classifier to further explore the models.

Even after applying these rigorous validation and evaluation techniques, the results continued to indicate 100% accuracy. As a team, a decision was made to choose a Decision Tree Classifier as the final model. Consequently, the top 80 features or symptoms identified by the Decision Tree Classifier were used in the final implementation of the chosen approach.

A screenshot of a computer program

Description automatically generated

Figure 4 Code snippet showing the implementation of decision tree classifier for feature importance.

**d. Data Visualization:**

Three different visualizations were developed to allow for insights into the dataset. The first one was a count plot showing the distribution of the different diseases across the dataset. This visualization provides a clear representation of the frequency or count of each disease in the dataset. In a balanced dataset, where all diseases have an equal number of instances, the count plot would display uniform counts for each disease category. However, if the dataset is imbalanced, the count plot would highlight the skewed distribution, with some diseases being overrepresented or underrepresented.

A graph with blue and white lines

Description automatically generated

Figure 5. Horizontal Bar Chart showing the distribution of symptoms across the different diseases.

Secondly, a correlation matrix, which displays the pairwise correlations between all the features (symptoms) in the dataset was visualised. This visualization helps identify highly correlated features, which can be useful for feature selection or detecting potential multicollinearity issues. The correlation matrix can also reveal interesting patterns or relationships between different symptoms.

A screen shot of a graph

Description automatically generated

Figure 6. Correlation Matrix showing the correlation between the different features in the dataset.

Lastly, a boxplot was created to search for outliers in the dataset. Boxplots are particularly useful for identifying outliers, which are data points that significantly deviate from the rest of the data. In the context of this medical diagnosis dataset, outliers could represent rare or unusual symptom combinations or potentially erroneous data points. The boxplot visualization allows for easy identification of these outliers by displaying them as individual points outside the whiskers of the plot. Detecting and handling outliers is crucial for ensuring the robustness and reliability of the machine learning models.

These pre-processing steps were essential for preparing the data for machine learning models by eliminating redundant or uninformative features, ultimately improving the model's overall performance, interpretability, and reliability.

A graph with many lines

Description automatically generated with medium confidence

Figure 7. Box plot for the detection of outliers. This figure shows there are no outliers.

**e. Details of Approach:**

Initially, we trained and evaluated several classification models, including logistic regression, decision trees, random forest, support vector machines (SVM), naive Bayes, and neural networks. Surprisingly, all models consistently achieved perfect scores across multiple evaluation metrics, including accuracy, precision, recall, and F1-score. While this outcome is rare and initially promising, it's essential to acknowledge that real-world datasets may present challenges that could impact the model’s performance differently.

Given the exceptional performance of individual models, we proceeded to explore the ensemble method stacking to leverage the strengths of different models, particularly decision trees and neural networks. Ensemble methods offer an opportunity to combine the predictive power of diverse algorithms, potentially leading to improved performance and generalization[15].

We implemented a stacking classifier, which combines the predictions of multiple base estimators, which were decision trees and neural networks in this project. Despite our high expectations, the stacking classifier also achieved perfect accuracy, indicating successful model integration (See **Figure 9**). While this is encouraging, it's crucial to interpret these results with caution and consider potential issues such as overfitting or data leakage. Overfitting can occur when the model learns the training data too well, including its noise and irregularities, leading to poor generalization on new data. Data leakage refers to the unintended exposure of information from the test set during training, resulting in an overly optimistic performance evaluation[19].

We acknowledge the limitations of our analysis, particularly the idealized nature of the dataset used for evaluation. Real-world medical datasets often contain noise, imbalance, or unseen patterns that can significantly impact model performance. Therefore, to best understand the performance of this model we’ve developed, practical applicability and generalization capabilities, we need to validate the model on external real-world medical datasets.

In conclusion, although our analysis has highlighted the remarkable performance of our ensemble method, it's essential to maintain a critical perspective and continue researching ways to refine our models to ensure their applicability to real-world scenarios.

**A screenshot of a computer program

Description automatically generated**Figure 8. Code snippet showing the implementation of Stacking classifier.

A screenshot of a computer

Description automatically generatedFigure 9. Results for model's performance and predictions using unseen dataset.

## 6. Contribution of Team Members

As a team of two members, we adopted a collaborative strategy where we both worked together on each section from start to finish. The initial phase involved extensive literature searches and dataset selection conducted by both of us, to gain a comprehensive understanding of the different approaches and methodologies related to the project.

For the coding phase, we split the task into two – Angel worked on implementing the initial preprocessing steps and data visualization techniques. While Millicent worked on the other sections including the validation and evaluation techniques to thoroughly explore the models' performance, particularly considering the perfect accuracy achieved after using the extracted features. For the feature extraction phase since two different approaches were used and each of us implemented a specific one: Angel worked on the use of Pearson correlation approach and variance thresholding approach while Millicent worked on the use of the decision tree classifier approach.

Following that, we collectively decided to employ the stacking method drawing upon our shared understanding. The final phase involved compiling the project report. We both took responsibility for writing up on the sections we worked on.

To ensure quality and consistency, we independently proofread the entire report before agreeing on the final version for submission.

## 7. Why your project is challenging

The project was challenging due to the implementation of advanced machine learning techniques, specifically the stacking ensemble method that integrated decision trees and neural networks. Carefully tuning and analysing the integration of these fundamentally different model architectures required a deep understanding of their respective strengths. Additionally, dealing with healthcare data necessitated ensuring high reliability and accuracy, as decisions based on this model are life critical. Navigating the use of right features for the model added complexity to ensure the model's suitability for use.

In summary, the high technical goals and strong dedication needed to overcome these challenges highlight how complex the project was.

## References

[1]

D. J. Park, M. W. Park, H. Lee, Y.-J. Kim, Y. Kim, and Y. H. Park, “Development of machine learning model for diagnostic disease prediction based on laboratory tests,” *Scientific Reports*, vol. 11, no. 1, Apr. 2021, doi: https://doi.org/10.1038/s41598-021-87171-5.

[2]

M.-A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. Yáñez-Márquez, and J.-C. Salgado-Ramírez, “Classification of Diseases Using Machine Learning Algorithms: A Comparative Study,” *Mathematics*, vol. 9, no. 15, p. 1817, Jul. 2021, doi: https://doi.org/10.3390/math9151817.

[3]

M. Ahsan and Z. Siddique, “Machine Learning Based Disease Diagnosis: A Comprehensive Review A Preprint Abbreviation ANN Artificial Neural Network CNN Convolutional Neural Network DL Deep Learning DNN Deep Neural Networks DT Decision Trees KMC K-Means Clustering KNN K-Nearest Neighbors LR Logistic Regression ML Machine Learning NN Neural Network RF Random Forest SVM Support Vector Machine,” 2022. Accessed: May 01, 2024. [Online]. Available: https://arxiv.org/pdf/2112.15538

[4]

I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Computer Science*, vol. 2, no. 3, pp. 1–21, Mar. 2021, doi: https://doi.org/10.1007/s42979-021-00592-x.

[5]

J. Han, J. Pei, and H. Tong, *Data Mining*. Morgan Kaufmann, 2022.

[6]

M. Badawy, “Healthcare predictive analytics using machine learning and deep learning techniques: a survey,” *ProQuest*, vol. 10, no. 40, p. 40, Dec. 2023, doi: https://doi.org/10.1186/s43067-023-00108-y.

[7]

The Health Informatics Service (THIS), “The Health Informatics Service,” *www.this.nhs.uk*, Jan. 11, 2021. https://www.this.nhs.uk/insights/article/predictive-analytics-an-emerging-asset-in-the-healthcare-industry

[8]

A. Sharma, A. Lysenko, S. Jia, K. A. Boroevich, and T. Tsunoda, “Advances in AI and machine learning for predictive medicine,” *Journal of Human Genetics*, pp. 1–11, Feb. 2024, doi: https://doi.org/10.1038/s10038-024-01231-y.

[9]

A. Alanazi, “Using machine learning for healthcare challenges and opportunities,” *Informatics in Medicine Unlocked*, vol. 30, p. 100924, 2022, doi: https://doi.org/10.1016/j.imu.2022.100924.

[10]

Enrico Barbierato and A. Gatti, “The Challenges of Machine Learning: A Critical Review,” *Electronics*, vol. 13, no. 2, pp. 416–416, Jan. 2024, doi: https://doi.org/10.3390/electronics13020416.

[11]

H. Hakkoum, I. Abnane, and A. Idri, “Interpretability in the medical field: A systematic mapping and review study,” *Applied Soft Computing*, vol. 117, p. 108391, Mar. 2022, doi: https://doi.org/10.1016/j.asoc.2021.108391.

[12]

S. Decherchi, E. Pedrini, M. Mordenti, A. Cavalli, and L. Sangiorgi, “Opportunities and Challenges for Machine Learning in Rare Diseases,” *Frontiers in Medicine*, vol. 8, Oct. 2021, doi: https://doi.org/10.3389/fmed.2021.747612.

[13]

I. Y. Chen, E. Pierson, S. Rose, S. Joshi, K. Ferryman, and M. Ghassemi, “Ethical Machine Learning in Healthcare,” *Annual Review of Biomedical Data Science*, vol. 4, no. 1, pp. 123–144, Jul. 2021, doi: https://doi.org/10.1146/annurev-biodatasci-092820-114757.

[14]

E. Tuba, I. Strumberger, T. Bezdan, N. Bacanin, and M. Tuba, “Classification and Feature Selection Method for Medical Datasets by Brain Storm Optimization Algorithm and Support Vector Machine,” *Procedia Computer Science*, vol. 162, no. 7, pp. 307–315, 2019, doi: https://doi.org/10.1016/j.procs.2019.11.289.

[15]

A. Mohammed and R. Kora, “A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, Feb. 2023, doi: https://doi.org/10.1016/j.jksuci.2023.01.014.

[16]

R. J. Ellis, R. M. Sander, and A. Limon, “Twelve key challenges in medical machine learning and solutions,” *Intelligence-Based Medicine*, p. 100068, Jun. 2022, doi: https://doi.org/10.1016/j.ibmed.2022.100068.

[17]

M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review,” *Healthcare*, vol. 10, no. 3, p. 541, Mar. 2022, doi: https://doi.org/10.3390/healthcare10030541.

[18]

J. Brownlee, “Stacking Ensemble Machine Learning With Python,” *Machine Learning Mastery*, Apr. 09, 2020. https://machinelearningmastery.com/stacking-ensemble-machine-learning-with-python/ (accessed May 5AD).

[19]

J. Brownlee, “Overfitting and Underfitting With Machine Learning Algorithms,” *Machine Learning Mastery*, Aug. 12, 2019. https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/

[20]

I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[21]

G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: https://doi.org/10.1016/j.compeleceng.2013.11.024.

[22]

Muhammad Summair Raza and U. Qamar, *Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications*. Springer, 2017.