People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

University Larbi Tébessi - Tébessa

Faculty of Exact Sciences and Natural and Life Sciences

Department: Mathematics and Computer Science

Final thesis

To obtain the MASTER diploma

Domain: Mathematics and Computer Science

Field: Computer Science

Specialty: Systems and Multimedia

A data-driven machine learning approach for migraine analysis and classification

Presented by:

Belghit Amel

In front of the jury:

Dr. Bendib Issam	MCA Larbi Tebessi University	President
Dr. Laifa Hassiba	MAB Larbi Tebessi University	Thesis Supervisor
Dr. Brik Assia	MAB Larbi Tebessi University	Examiner

College year: 2023/2024

بسم الله الرحمن الرحيم

Abstract

Migraine is a severe neurological condition characterized by intense headaches and accompanying symptoms, impacting millions of people worldwide. Accurately categorizing migraine types is crucial for effective diagnosis and treatment. Traditional classification methods often rely on subjective clinical judgment, leading to inconsistencies. This thesis explores advanced machine learning and deep learning techniques to improve the precision of migraine classification using a comprehensive dataset. The research comprises two main approaches. The first approach utilizes ensemble learning algorithms, particularly XGBoost, to classify migraine types based on diverse features. This method combines descriptive, diagnostic, and predictive analyses to offer a comprehensive understanding of the dataset. Extensive experimentation demonstrates that XGBoost performs better than traditional machine learning models in terms of accuracy and efficiency. The second approach introduces a unique method for converting tabular data into image format, enabling the use of Convolutional Neural Networks (CNNs) for classification. This innovative technique harnesses the powerful feature extraction capabilities of CNNs, resulting in superior classification accuracy compared to traditional and ensemble learning models. The integration of these advanced methodologies significantly enhances the accuracy and reliability of migraine classification. The implications of these findings are substantial for medical informatics and migraine management, enabling more targeted and effective treatment plans. Moreover, the novel methodologies presented in this thesis can be adapted to other medical classification issues, further extending their potential impact. Overall, this research demonstrates the effectiveness of advanced machine learning and deep learning techniques in enhancing the classification of migraine types, laying the groundwork for more precise and reliable diagnostic tools in healthcare.

Keywords: Migraine, Artificial Intelligence, Machine learning, XGBoost, CNN. Intelligent-methods.

Résumé

La migraine est une maladie neurologique grave caractérisée par des maux de tête intenses et des symptômes qui l'accompagnent, touchant des millions de personnes dans le monde. Catégoriser avec précision les types de migraine est crucial pour un diagnostic et un traitement efficaces. Les méthodes de classification traditionnelles reposent souvent sur un jugement clinique subjectif, ce qui entraîne des incohérences. Cette thèse explore des techniques avancées d'apprentissage automatique et d'apprentissage profond pour améliorer la précision de la classification de la migraine à l'aide d'un ensemble de données complet. La recherche comprend deux approches principales. La première approche utilise des algorithmes d'apprentissage d'ensemble, en particulier XGBoost, pour classer les types de migraine en fonction de diverses caractéristiques. Cette méthode combine des analyses descriptives, diagnostiques et prédictives pour offrir une compréhension complète de l'ensemble de données. Une expérimentation approfondie démontre que XGBoost est plus performant que les modèles d'apprentissage automatique traditionnels en termes de précision et d'efficacité. La deuxième approche introduit une méthode unique pour convertir des données tabulaires en format d'image, permettant l'utilisation de réseaux de neurones convolutifs (CNN) pour la classification. Cette technique innovante exploite les puissantes capacités d'extraction de caractéristiques des CNN, ce qui se traduit par une précision de classification supérieure par rapport aux modèles d'apprentissage traditionnels et d'ensemble. L'intégration de ces méthodologies avancées améliore considérablement la précision et la fiabilité de la classification de la migraine. Les implications de ces résultats sont considérables pour l'informatique médicale et la gestion de la migraine, permettant des plans de traitement plus ciblés et plus efficaces. De plus, les nouvelles méthodologies présentées dans cette thèse peuvent être adaptées à d'autres problèmes de classification médicale, élargissant ainsi leur impact potentiel. Dans l'ensemble, cette recherche démontre l'efficacité des techniques avancées d'apprentissage automatique et d'apprentissage profond pour améliorer la classification des types de migraine, jetant ainsi les bases d'outils de diagnostic plus précis et plus fiables dans le domaine des soins de santé.

Mots clés : Apprentissage Automatique, Classification, la migraine, Intelligence artificiel, XGBoost, CNN, deep learning

ملخص:

الصداع النصفي هو حالة عصبية حادة تتميز بصداع شديد وأعراض مصاحبة، وتؤثر على ملايين الأشخاص في جميع أنحاء العالم. يعد التصنيف الدقيق لأنواع الصداع النصفي أمرًا بالغ الأهمية للتشخيص والعلاج الفعال. تعتمد طرق التصنيف التقليدية في كثير من الأحيان على الحكم السريري الشخصي، مما يؤدي إلى التناقضات. تستكشف هذه الأطروحة التعلم الآلي المتقدم وتقنيات التعلم العميق لتحسين دقة تصنيف الصداع النصفي باستخدام مجموعة بيانات شاملة. يتكون البحث من نهجين رئيسيين. يستخدم النهج الأول خوارزميات التعلم الجماعي، وخاصة XGBoost، لتصنيف أنواع الصداع النصفي على أساس ميزات متنوعة. تجمع هذه الطريقة بين التحليلات الوصفية والتشخيصية والتنبؤية لتقديم فهم شامل لمجموعة البيانات. توضح التجارب المكثفة أن أداء XGBoost أفضل من نماذج التعلم الآلي التقليدية من حيث الدقة والكفاءة. يقدم النهج الثاني طريقة فريدة لتحويل البيانات الجدولية إلى تنسيق صور، مما يتيح استخدام الشبكات العصبية التلافيفية ((CNNsللتصنيف. تستغل هذه التقنية المبتكرة إمكانات استخلاص الميزات القوية لشبكات CNN، مما يؤدي إلى دقة تصنيف فائقة مقارنة بنماذج التعلم التقليدية والمجمّعة. إن دمج هذه المنهجيات المتقدمة يعزز بشكل كبير دقة وموثوقية تصنيف الصداع النصفي. الآثار المترتبة على هذه النتائج كبيرة بالنسبة للمعلوماتية الطبية وإدارة الصداع النصفي، مما يتيح خطط علاج أكثر استهدافا وفعالية. علاوة على ذلك، يمكن تكييف المنهجيات الجديدة المقدمة في هذه الأطروحة مع قضايا التصنيف الطبي الأخرى، مما يزيد من تأثيرها المحتمل. بشكل عام، يوضح هذا البحث فعالية التعلم الآلي المتقدم وتقنيات التعلم العميق في تعزيز تصنيف أنواع الصداع النصفي، ووضع الأساس لأدوات تشخيصية أكثر دقة وموثوقية في مجال الرعاية الصحية.

الكلمات المفتاحية :الذكاء الاصطناعي، الطرق الذكية، الصداع النصفي

Thank

First of all, I would like to express my gratitude to the editors of this dissertation, Miss. Hassiba laifa, for their patience, availability, and, above all, their sound advice. Their understanding was particularly valuable, as it enabled us to make significant progress and perform exceptionally well throughout these months.

I would also like to thank the members of the jury:

- Dr. Bendib Issam
- Dr. Brik Assia

I would also like to thank all the students with whom I have been happy to study during these years of university.

I thank all the teachers that I have been able to meet and benefit from.

I would like to thank my colleagues in the Mathematics and Computer Science department and all those who have helped me directly or indirectly in the realization and accomplishment of this work.

Dedication

I am grateful to Allah for blessing me with the strength and courage to complete this research. Throughout the process, I had many people who supported and guided me through every obstacle that I faced. Therefore, I would like to express my deepest appreciation to those involved in this thesis.

Firstly, I would like to express my gratitude and heartfelt thanks to my subject teacher, Mrs. Hassiba Laifa, who provided me with ideas and guidance from the beginning till the end.

Last but not least, I would like to extend my appreciation to my beloved parents who have been my source of inspiration and strength throughout this research. Their continued moral, spiritual, emotional, and financial support helped me to complete this thesis.

I would also like to thank my wonderful sisters Ilhem, Chaima, and Djihen.

Also my dear siblings Akram, Ayoub, and Yaakoub.

And all my friends who shared their ideas and helped make my research better. I hope we can achieve a good grade for our effort.

For those who have not been mentioned here, I would like to thank them for their help, friendship, and countless support. I am grateful for their kindness and support. May Allah bless them all.

Amel

Table of Contents

List of Figures

List of tables

List of abbreviations

LR Linear Regression

DT Decision Tree

RF Random Forest

SVM Support Vector Machine

ANN Artificial Neural Network

KNN K-nearest Neighbour

GBRT Gradient boosting regression tree

XGBoost eXtreme Gradient Boosting

DBSCAN Density-Based Spatial Clustering of Applications with Noise

FHM familial hemiplegic

MWA Migraine with Aura

SHM Sporadic hemiplegic Migraine

TTH Tension-type Headache

OTIs Diffusion tensor images

MOH Medication overuse headache

General Introduction

Migraine is a prevalent and debilitating neurological disorder affecting millions worldwide. Accurate classification of migraine types is crucial for appropriate treatment and management. Traditional methods of classification rely heavily on clinical judgment, which can be subjective and inconsistent. This thesis explores advanced machine learning and deep learning techniques to improve the classification accuracy of migraine types using a comprehensive dataset.

In this research, we aim to tackle the complexity and variability in migraine symptoms that make classification challenging. We will use a rich dataset containing various migraine types and apply sophisticated data-driven approaches to enhance the precision of classification. Specifically, we will investigate ensemble learning algorithms, with a focus on XGBoost, and introduce a novel method to transform tabular data into images for application in Convolutional Neural Networks (CNNs). The integration of descriptive, diagnostic, and predictive analyses will provide a holistic view of the data, improving both the accuracy and reliability of migraine-type classification.

Problematic

Migraine classification presents several challenges. The primary issues include:

Subjectivity and Inconsistency in Clinical Diagnosis: Traditional diagnostic methods are often subjective, leading to inconsistent results and misclassifications. Variability in symptoms further complicates the diagnosis, making it difficult to standardize the classification process.

Complexity and Variability of Symptoms: Migraine symptoms can vary significantly between individuals, leading to overlapping characteristics among different migraine types. This variability complicates the classification process, requiring more sophisticated techniques to distinguish between types accurately.

Need for Robust Data-Driven Methods: There is a pressing need for robust, data-driven methods that can handle the complexity of migraine classification. Traditional machine learning models have shown promise, but they often fall short in terms of accuracy and reliability. Advanced techniques, such as ensemble learning and deep learning, offer potential solutions but require further exploration and validation.

Research Objectives

1. To investigate the potential of machine learning and deep learning algorithms in the classification of migraine types.
2. To develop a robust framework for descriptive, diagnostic, and predictive analysis of migraine data.
3. To propose and evaluate novel methods for transforming tabular data into a format suitable for deep learning applications, specifically Convolutional Neural Networks (CNNs).

Propositions

1. **Proposition 1**: Implementing ensemble learning algorithms, particularly XGBoost, to classify migraine types based on a rich dataset. This approach will be evaluated against traditional machine learning models to establish its efficacy.
2. **Proposition 2**: Develop a novel method to transform tabular migraine data into images, enabling the application of CNNs for classification. This method aims to harness the feature extraction capabilities of CNNs, potentially outperforming traditional and ensemble learning models.

This thesis is structured into several key chapters:

- **Chapter 1**: Basic Concepts - An overview of data analysis, machine learning, and deep learning.
- **Chapter 2**: Literature Review - A review of existing works on migraine classification using machine learning, highlighting gaps and opportunities.
- **Chapter 3**: Proposed Method - Detailed methodologies for the ensemble learning approach and the tabular-to-image transformation method with CNNs.
- **Results and Discussion**: Presentation and interpretation of the findings from the implemented methods.
- **Conclusion**: Summarizing the contributions, implications, and potential future work.

Chapter I: Basic Concepts

Introduction

Data analysis and machine learning are cutting-edge technologies that analyze data to derive valuable insights and enable predictive capabilities. Data analytics involves examining large datasets to uncover patterns and trends, while machine learning focuses on developing algorithms that learn from data to make predictions or decisions autonomously. Together, these technologies drive innovation and efficiency by harnessing the power of data to inform strategic initiatives and improve decision-making processes.

In this chapter, we will cover the basic concepts of using data analysis in healthcare. The chapter is divided into three sections. The first section covers preliminary concepts related to data analysis, including its definition and types. In the second section, we delve into Machine Learning, defining it, discussing the two types of learning, and exploring research algorithms. Lastly, the third section focuses on Data Analysis in Healthcare, where we discuss the Nature of Healthcare Data, Applications of Big Data in Healthcare, and provide examples of Big Data Analysis for Healthcare.

1. Data Analysis

In this section, we introduce data analysis as a fundamental concept related to our study, along with its various types.

1.2 Definition:

Data analysis is a very interdisciplinary field that has adopted aspects from many other scientific disciplines such as statistics, machine learning, pattern recognition, system theory, operations research, and artificial intelligence [1]. As a multidisciplinary concept, data analysis is defined as the means to acquire data from diverse sources, process them to elicit meaningful patterns and insights, and distribute the results to proper stakeholders [2]. To summarize, the primary objective of data analysis is to get valuable information from the data and make judgments using this knowledge.

Data analysis is concerned with the extraction of actionable knowledge and insights from data [3]. Tabular data are pervasive. Although tables often describe multivariate data without explicit network semantics, exploring the data modeled as a graph or network for analysis may be advantageous. Even when a given table design conveys some static network semantics, analysts may want to look at multiple networks from different perspectives, at different levels of abstraction, and with different edge semantics. Tabular data come in many forms, each unique in its schematic and semantic structure depending on the technology used and the data owner’s goal. The term “tabular data” is thus fairly broad and can be interpreted as multivariate data or attribute relationship graphs. [4]

Tabular data—in contrast to image or language data—are heterogeneous, leading to dense numerical and sparse categorical features. Furthermore, the correlation among the features is weaker than the one introduced through spatial or semantic relationships in image or speech data.

Heterogeneous data are the most commonly used form of data. It is ubiquitous in many crucial applications, such as medical diagnosis based on patient history, predictive analytics for financial applications (e.g., risk analysis, estimation of creditworthiness, the recommendation of investment strategies, and portfolio management), click-through rate (CTR) prediction, user recommendation systems, customer churn prediction, cybersecurity, fraud detection, psychology,

anomaly detection, and so forth. In all these applications, a boost in predictive performance and robustness may benefit end users and companies that provide such solutions. Simultaneously, this requires handling many data-related pitfalls, such as noise, impreciseness, different attribute types, and value ranges, or the missing value problem and privacy issues. [5]

2. Types of Data Analysis

The primary goal of the study understanding the past, examining the present, or forecasting the future determines the sort of analysis that is used. Descriptive, diagnostic, predictive, and prescriptive analytics, which are the four most popular forms of analytics, are interconnected tools that assist companies in making the most of their available data. These several analysis types each provide unique insights. The four categories of analytics are discussed in this section: prescriptive, predictive, diagnostic, and descriptive analytics.

2.1 Descriptive Analysis

Descriptive analysis, also called business reporting, uses the data to answer the question of “what happened and/or what is happening?”. [6] Data gathered is organized as bar charts, graphs, pie charts, maps, scatter diagrams, etc., for easy visualization which gives insight into what the data implies. This form of data presentation is often called a dashboard, mimicking the dashboard of a car which gives information on speed, engine status, petrol left in the tank, distance traveled, etc. A typical example of descriptive analytics is the presentation of population census data which classifies population across a country by sex, age groups, education, income, population density, and similar parameters.[7]

2.2 Diagnostic Analysis

The diagnostic analysis is also considered as an advanced form of data analysis. It helps to examine insight of data that answers the question “Why did it happen?”. It considers different attributes and features information to discover the relations. It is likewise known as data discovery, data mining, and correlation techniques. It helps to analyze insight of data and attempts to interpret the causes of events and behaviors. In health care, diagnostic analytics explore the information and make correlations using different attribute information [8]. This type of analysis is often used in the field of medicine to identify the relationship between causes and effects.

2.3 Predictive Analysis

Predictive analysis which is the “forecasting or extrapolation” phase incorporates the descriptive analytics output as well as some machine learning (ML) algorithms and simulation techniques to build accurate models that predict the future. It answers the questions “What will happen?” and “Why will it happen?” in the future. [9]

Predictive analysis delivers a significant return on investment compared to other BI technologies. Analytical models can optimize a wide variety of business processes in different industries and functional areas. [10] Predictive analysis includes a set of statistical models, machine learning algorithms, and data mining tools that analyze historical data and real-time data to predict future events. It includes two components:

1) Predictive models designed to predict new/future observations or scenarios.

2) Methods for evaluating a model's predictive performance (predictive accuracy).

Algorithm-based methods such as supervised machine learning models (regression, classification) and unsupervised machine learning models (clustering, association) are commonly used in predictive analytics.

2.4 Prescriptive Analysis

Prescriptive analysis which is the “recommender or guidance” phase provides enterprises with adaptive, automated, time-dependent, and optimal decisions. Its goal is to bring business value through better strategic and operational decisions. In general, prescriptive analytics is predictive analytics that prescribes one or more courses of action and shows the likely outcome/influence of each action. It answers the questions “What should I do?” and “Why should I do it?”. It is purely built on the “what-if” scenarios. The main elements of prescriptive analytics are optimization, simulation, and evaluation methods. Simply put, it provides advice based on predictions and enterprise constraints. [11]

The prescriptive analysis makes the users prescribe the possible outcomes to attain a better solution or it provides better advice. This attempts to enumerate future decisions before plunging into some decisions. These analytics not only predict what is about to happen but also why and how the problem exists provide possible recommendations based upon the actions and provide a better recommendation of predictions. These analytics are used to predict multiple future predictions and make the company use the possible outcomes for their future. [12]

Such types can be summed up as follows: Both diagnostic and descriptive analysis use historical evidence to understand what occurred and why. Both prescriptive and predictive analytics make use of previous data to forecast future events and suggest courses of action for improving those results. Figure 1 summarizes the four types of analysis.

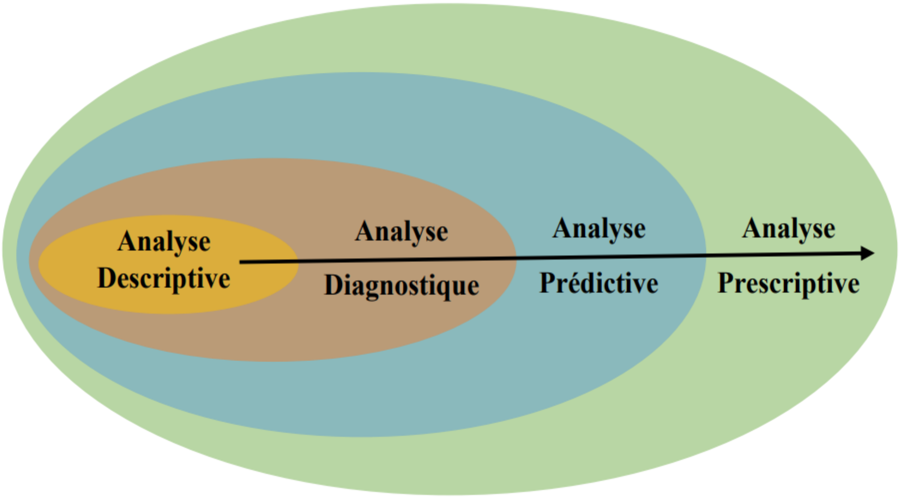


Figure 01: Types of data analysis.

3. Machine learning

This section introduces the concept of Machine Learning and explores its various types.

3.1. Definition

Machine learning is a very active sub-field of artificial intelligence concerned with the development of computational models of learning. [13] Arthur Samuel in his seminal work defined machine learning as “a field of study that gives computers the ability to learn without being explicitly programmed”. [14] A learning problem can be defined as the problem of improving some measure of performance when executing some tasks, through some type of training experience. [15] Machine learning methods are divided into several categories based on the task's purpose.

3.2. Machine learning classes

Machine learning is divided into two main categories depending on the type of dataset used for learning: supervised learning and unsupervised learning.

3.2.1. Supervised learning

Supervised learning is the most used in machine learning algorithms. It is defined by finding a function f that maps an input x to its output y based on examples to learn. The input for training is presented with a pair of examples that contain a set of features (X1, X2, …, Xn) and their desired target Y, meaning the data is labeled, and each input to its known output Y before starting training. Two different problems require the use of supervised learning: classification and regression.

- **Classification**

Which means that each input X belongs to a predefined group or class Y. In this case, the supervised task assumes that Y is of the categorical type (two or more classes).

- **Regression**

In the field of regression prediction, whether from traditional mathematical statistics methods or machine learning algorithms in recent years, researchers at home and abroad have proposed many research methods for prediction [16] Mathematically, the result must be of a continuous numerical value (data with real values representing a quantity for example: salary, age, time, etc.).

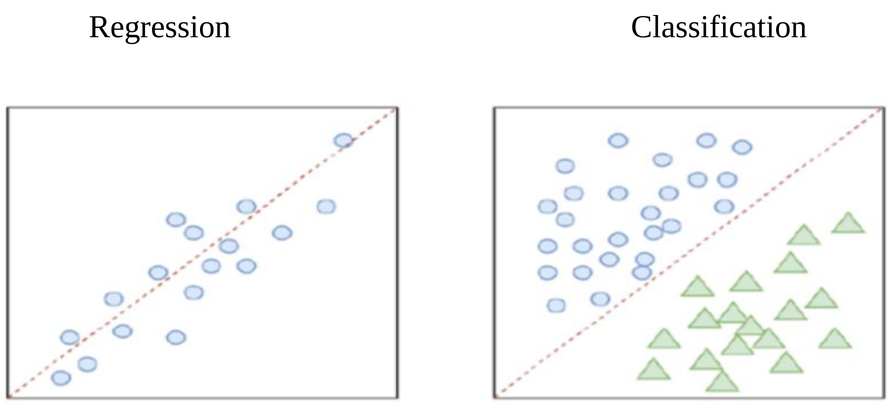


Figure 02: Supervised learning.

- **Supervised Machine Learning Algorithms**

In this section, we cite commonly used supervised machine learning models.

- **Linear Regression:** One of the maximum general, extensive statistical and ML algorithms is linear regression. It's applied to identify linear relationships within one or farther identifiers. Two types of linear regression: multiple regressions (MLR) and simple regression. Different researchers are researching polynomial, and linear regression yet compare their effectiveness using the accession to optimistic precision and prediction.

Linear regression is the mathematical experiment conducted for quantifying and evaluating the familiarity of the calculated attributes. Therefore, sketchy regression and correlation are experiments where a boffin in perception the bonding into two attributes to count the influence of disorders. Linear regression is generally conducted in mathematical techniques. It's likely to identify the prediction model yet affect the versus multiplex input attributes.

The methods utilized, whether for regression or classification, are all aimed at locating the right answer; the only differences are in how the results are shown and how they are discovered.

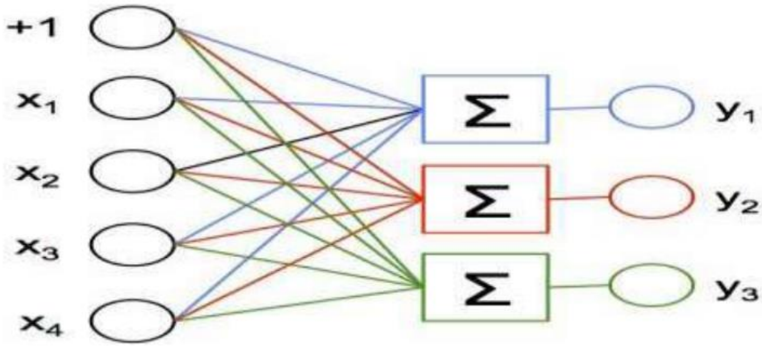


Figure 03: Architectural model of multi-variation linear regression.

- **Logistic Regression:** Logistic regression has broadly applied to different fields of the experiment, such as hygiene science, to learn the risk factor attached to the illness. Some surveys depending on Health Survey and Demographic are made predicting mixed modeling i.e., multistage sampling, stratified and probabilistic with unbalanced magnification in the survey. These compound diagrams must hold to calculate faithful outcomes. Although it is a relevant general issue and not well analyzed in the literature [29]. It is the preferential probabilistic structure. This structure generates inferior probability formation P(Y|X), (Y = destination variable and X = features). Given X, return to a probability formation over Y. Figure 04 represents an architecture model of the logistic regression. The outcome of the sigmoid function is explained as the probability of individual samples including the positive class, in the binary classification crux. An example, ∅ Z =P y = 1 x; w (Z = linear combination of the weights and the samples features Z = wTx. This algorithm is broadly used for classification. This algorithm is broadly used for classification. [17]

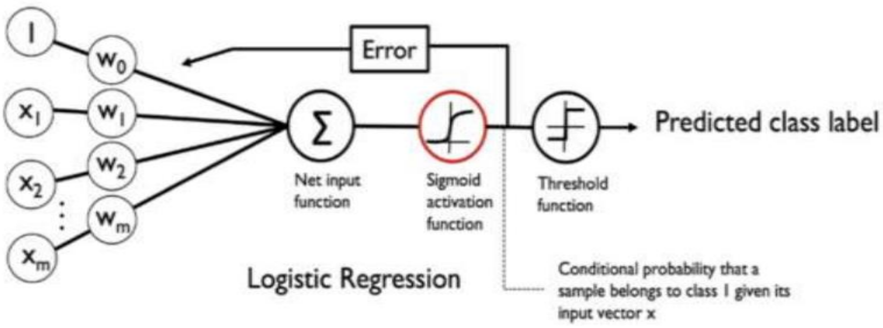


Figure 04: Architecture of a Logistic Regression Model.

- **Decision Tree (DT):** DT is one of the significant technologies in machine learning. Several sectors applied the Decision Tree algorithm and applied it in several applications. Decision Tree has three different algorithms that are C4.5, CART, and ID3. The ramification is the motive of offering objects to the class, which has various applications. A normal tree comprises roots, leaves, and branches. It's a predictive structure applied in machine learning, data mining, and statistics. In tree models, the destination variable can get a limited set of entities that are defined as classification trees; in the tree model, leaves represent the label of branches, and class represents the joins of fertilities that conduct with those labels of class. Decision trees can be constructed comparatively faster than any other method of classification. A decision tree is similar to the tree. To construct a tree, use the CART (Classification and Regression Tree algorithm). The structure of the decision tree:

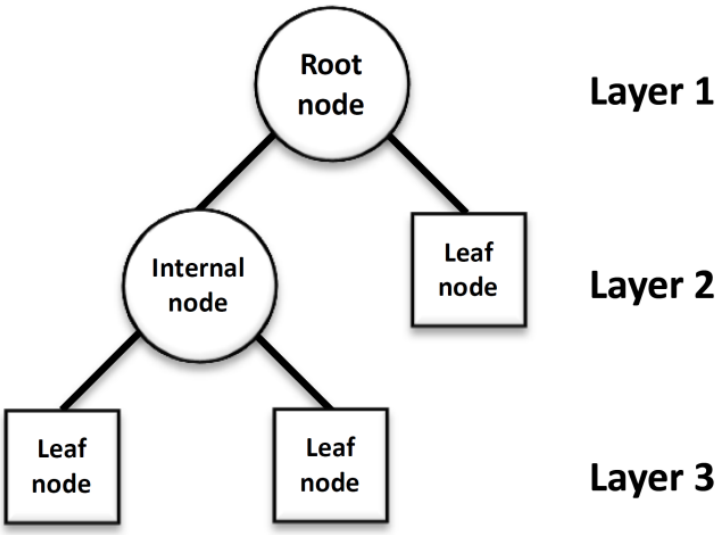


Figure 05: structure of the decision tree.

- **Random Forest (RF): The** Random Forest algorithm is an assembled method which combines the outcomes of various randomly built classification trees. Two elements of randomness are proposed for the building of the several trees. At First, every tree is built using the random bootstrapped form of the training dataset. Prediction contracts for unobserved datasets by getting a majority view of the individual trees. Random Forest packages in Python are used for implementation. In RF constructing a tree uses a random dataset. [17]

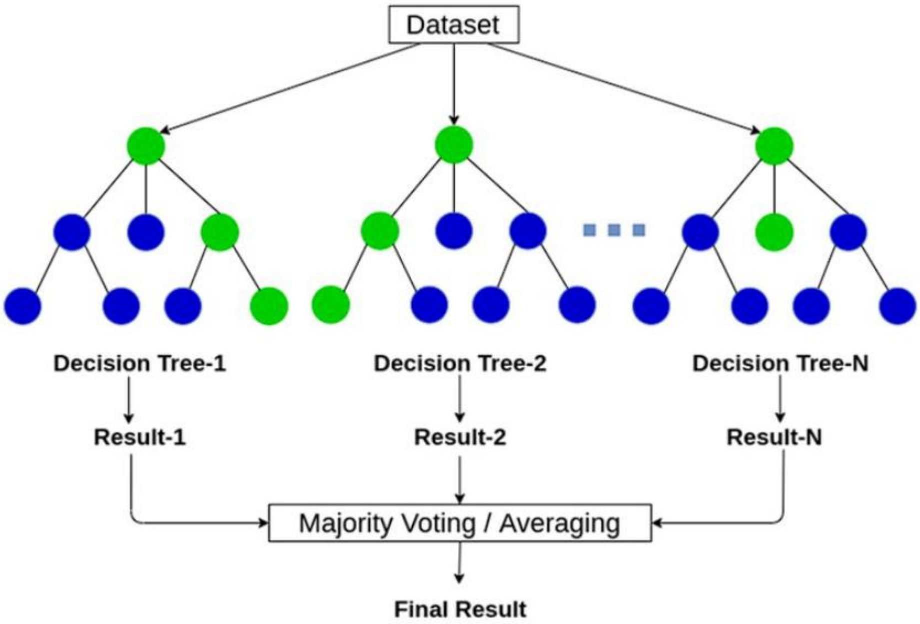


Figure 06: Random forest Architecture.

- **Support Vector Machine (SVM):** SVM is the significant ML algorithm usually developed in pattern recognition problems, classifying the image processing and network traffic for recognition. Too much research is working on skilling to develop Quality of Service (QoS) and indemnity aspects. Recent work in this sector has been solved by SVM. It acts more finely than any other classification network traffic for normalizing the difficulty. This research represents the aspect of SVM, its applications, and its concepts overview [18]. SVM is the strong learning method applied in binary classification. The SVM principal task is to search for the greatest hyperplane that can differentiate data properly into twice classes. Nowadays, multi-class classification is gained by mixing multiple binary support vector machines. [17]

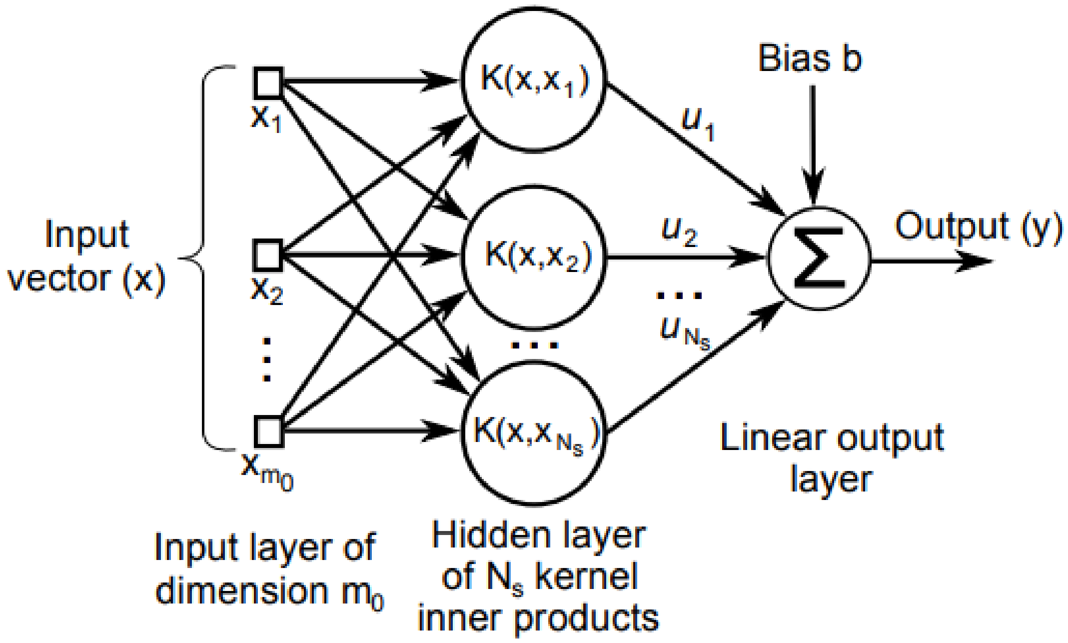


Figure 07: Architecture of support vector machine. [17]

- **Artificial Neural Network (ANN):** Recently, a multitude of methodologies and ideas from individual disciplinarian fields have increased in the particularly attractive research field of ANN. A neuron is a basic unit for making the nervous networks that perform communication and computational methods. The ANN is the working repetition of the facilitated method of the biological neuron. Yet, the aim is to reconstruct knowing data appraisal methods like classification, generalization, and pattern recognition using simple distributed and robust processing units named Processing Elements (PE) or artificial neurons. A chief benefit of the ANN access is that the ambit learning gives shares in neurons. Data processing is brought to pass in a collateral-distributed manner. ANNs are exceptionally collateral data processing instruments able to learn the working dependencies of the dataset. They have to be able to clearly categorize a high non-linear bearing yet, once trained, can categorize fresh datasets so much more swiftly than it should be probable by proving the structure logically. ANN formation is based on artificial neurons. [17]

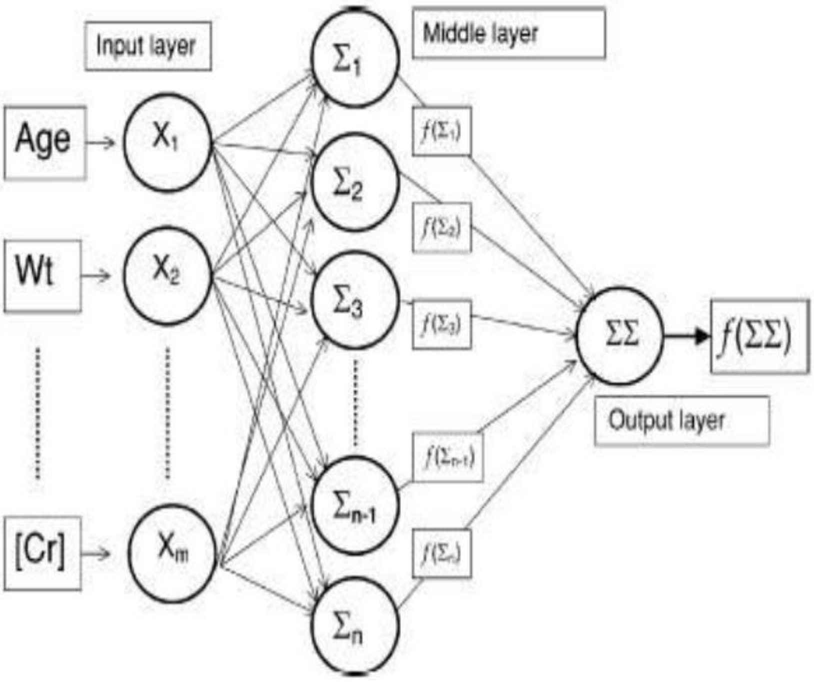


Figure 08: Every artificial neuron features a system node ('body') delineated by circles within the figure likewise as affiliations from ('dendrites') and affiliations to ('axons') alternative neurons that are shown as arrows sign in figure 08. In the last step, the output neuron gets the weighted aggregate of inputs and dispenses the non-linear functionality to the weighted aggregate. The results of this function make the outcome for the complete ANN. [17]

- **K-nearest Neighbour (KNN):** Distance-basis algorithms are broadly used for dataset classification difficulty. The KNN classification is the most exoteric distance-basis algorithm. Euclidean distance by this behavior of different datasets requires private resemblance measurement accommodated to the dataset features. Alignment is the SML system that graphs the input dataset, defining classes/groups. The main principle argument for investing an arrangement rein is that all of the dataset aims would be engaged to the groups and that every entity object would be engaged to a single group. Here, K is used as an amount of the closest neighbors in KNN.

- **Gradient boosting regression tree (GBRT):** Gradient boosting is an ensemble method that combines many weak learners such as the decision tree. As its name implies, a GBRT is a combination of gradient boosting and regression trees for regression problems that use ensembles of regression trees to reduce the error over a large single-tree model. Some important parameters in the GBRT model can generally be divided into two categories. One is about the frame of gradient boosting. For example, n estimators mean the number of regression trees. Loss is a given loss function for a regression tree in each stage to fit on its negative gradient. It determines how to compute the error values such as the least squares regression, least absolute deviation, and quantile regression. The learning rate controls the contribution of each decision tree, and there is a trade-off between the learning rate and n estimators. The other is about the decision tree. For example, max_depth limits the number of depths in each regression tree. Criterion is the function to calculate the quality of a split. Min_samples_leaf is the minimum number of instances required to be at a leaf node. Min_samples_split is used to determine whether an internal node can be split. [18]

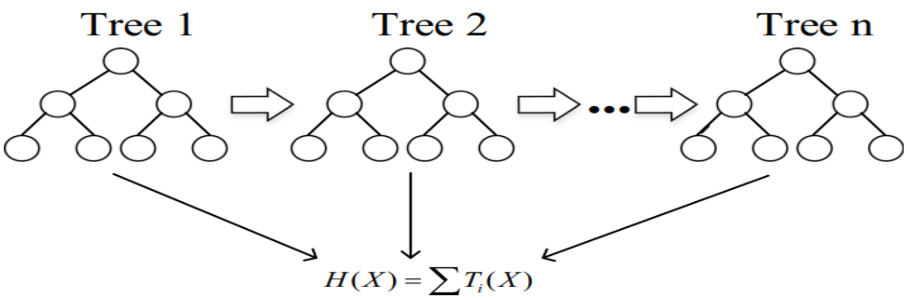


Figure 09: A GBRT model. Each tree is a decision tree. The tree Ti is built after tree T_{i-1}. The predicted value of GBRT is the sum of values that each tree predicts. The target input value of tree ti is the residual between the current predicted value and target value that can be defined as input i=$\sum_1^{i-1} Tj - Ytrue$ where Tj is the predicted result of the jth decision tree ($1 \leq j \leq i$ _1) and Y true is the target truth value of instances.[18]

- **Supervised machine learning algorithm comparison**

The following table, which compares a few supervised learning algorithms, provides an overview of all the algorithms mentioned above:

Algorithm	Algorithm type	advantages	disadvantages
Linear regression	Regression	-Quick training time, simple to understand and use.	-Performs optimally in simple linear connections.
Logistic Regression	Classification	Simple, interpretable, handles binary classification well	Assumes linear decision boundary, not suitable for non-linear data

Decision Trees	Classification/ Regression	Handles non-linear relationships, interpretable	Prone to overfitting, sensitive to small variations in data
Random Forest	Ensemble(Ensemble of decision trees, combines their predictions)	Robust against overfitting, handles large datasets	Less interpretable compared to individual decision trees
Support Vector Machines	Classification/ Regression	Effective in high-dimensional spaces	Computationally intensive for large datasets, requires careful tuning
Gradient Boosting Regression Tree	Ensemble (ensemble method that builds trees sequentially, correcting errors of previous trees)	High predictive accuracy, handles complex relationships	Prone to overfitting if not properly tuned, computationally intensive
Neural Networks	Classification/Regression	Handles complex relationships, state-of-the-art performance	Computationally intensive, requires large amounts of data
K-Nearest Neighbors	Classification /Regression	easy to understand, no training required	Computationally expensive for large datasets, sensitive to noise

Table 01: compares a supervised learning algorithm.

3.2.2. Unsupervised learning

Called unsupervised learning because unlike supervised learning above there are no correct answers and there is no teacher. Algorithms are left to their own devices to discover and present an interesting structure in the data. The unsupervised learning algorithms learn a few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction. [19] Unsupervised ML has many applications such as feature learning, data clustering, dimensionality reduction, anomaly detection, etc. In particular, recent unsupervised ML advances—such as the development of "deep learning" techniques have however significantly advanced the ML state of the art by facilitating the processing of raw data without requiring careful engineering and domain expertise for feature crafting [20].

The goal of applying clustering methods is to identify relevant subgroups in a given dataset without having a predefined hypothesis about what properties the subgroups might have.

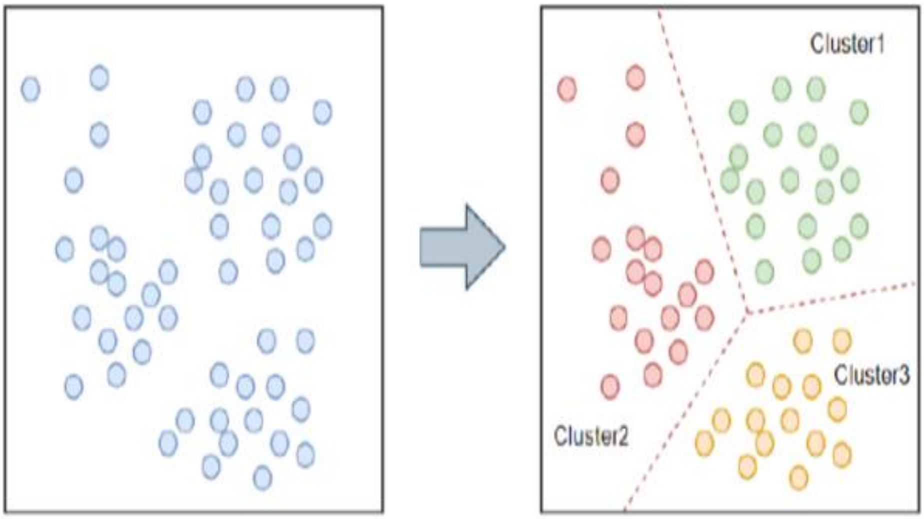


Figure 10: Unsupervised learning.

- **Unsupervised Machine Learning Algorithms**

- **K-Means algorithm:** The K-Means algorithm is a popular unsupervised machine learning method for clustering data. It aims to partition a given dataset into K distinct clusters, where each data point belongs to the cluster with the closest mean value. The algorithm iteratively updates the centroids of the clusters until convergence. According to Mitchell, T. M. (1997), K-Means is based on minimizing the within-cluster variance. On the other hand, Lloyd, S. P. (1982) introduced the original algorithm that iteratively assigns data points to their closest cluster centroid. The K-Means algorithm is a commonly used clustering technique. It aims to partition a dataset into K clusters, where each data point belongs to the cluster with the nearest mean. The algorithm follows a simple formula:

  - Initialize K cluster centroids randomly.

  - Assign each data point to the cluster with the closest centroid based on Euclidean distance.

  - Update the centroids by computing the mean of all data points in each cluster.

  - Repeat steps 2 and 3 until convergence when the cluster assignments no longer change significantly. [21]

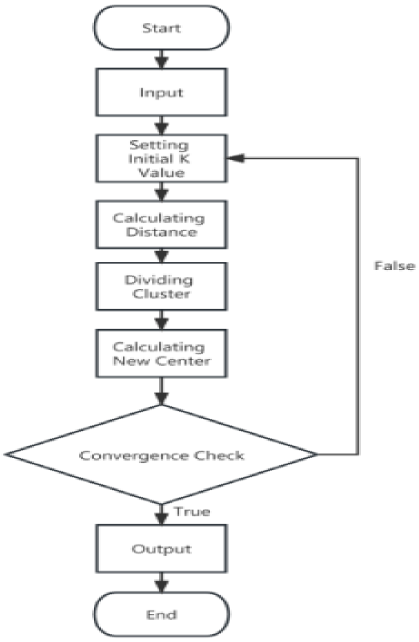


Figure 11: K-Means Algorithm Process. [21]

- **DBSCAN algorithm:** Density-Based Spatial Clustering of Application with Noise (DBSCAN) is a density-based clustering method. The basic idea of density-based clustering is to form a cluster that is dense enough and separated by sparse or low-density regions. The eps (epsilon) is a specified radius and density is estimated by calculating the number within eps-neighborhood of a point. The main concept of the DBSCAN algorithm is that each point in the cluster must contain at least Minpts number of points including itself within its eps-neighborhood. DBSCAN begins by randomly selecting a point, if it contains neighboring points fewer than the threshold (Minpts), it is labeled as noise temporarily; otherwise, it forms a cluster. Then the points within the eps-neighborhood of the selected point are added to the cluster and the cluster starts to expand**.** [22]

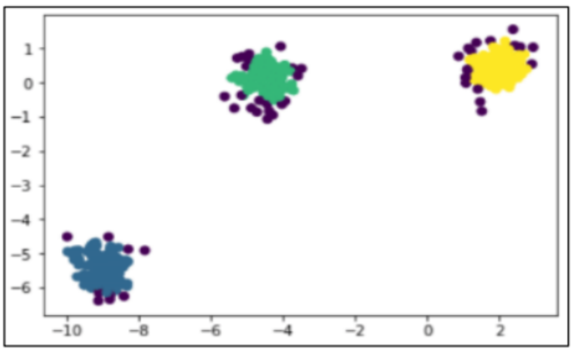


Figure 12: DBSCAN algorithm.

- **Hierarchical Clustering algorithm:** Hierarchical clustering is a method of cluster analysis in data mining that creates a hierarchical representation of the clusters in a dataset. The method starts by treating each data point as a separate cluster and then iteratively combines the closest clusters until a stopping criterion is reached. The result of hierarchical clustering is a tree-like structure, called a dendrogram, which illustrates the hierarchical relationships among the clusters.[23]

  - **Algorithm comparison of unsupervised learning**

Below is a table that compares various unsupervised learning algorithms:

Algorithm	Algorithm type	Advantages	disadvantages
K-means	Partitioning Clustering	- Works well with big datasets. - Simple to comprehend and apply	-The first choice of cluster centers determines performance. - Unsuitable for data with changing densities or irregular shapes
DBSCAN	Density Clustering	- Able to identify clusters of any shape - Does not require a prior specification of the number of clusters	-Is not appropriate for high-dimensional datasets - The choice of minimum density and epsilon determines performance.
Hierarchical Clustering	Hierarchical Clustering	- Able to show the clusters' hierarchical structure	-Large datasets are not appropriate for it because of its high time complexity

Table 02: compares unsupervised learning algorithms.

4. Data Analysis in Healthcare

The field of healthcare has seen a rapid increase in the applications of data analysis during the last decades. By utilizing different data analysis solutions, healthcare areas such as medical image analysis, disease recognition, outbreak monitoring, and clinical decision support have been automated to various degrees. Consequently, the intersection of healthcare and data analysis has received scientific attention to the point of numerous secondary studies [24].

The healthcare system aims to treat patients efficiently and safely through various approaches, including evidence-based decision support systems, diagnostic clinical genomics, and digital health. Health data is crucial for record keeping, clinical decision support, and health system management. Intelligent health data analytics methods are gaining utility, enabling predictive, personalized insights for clinical and health system management.[25]

In other ways, Data analysis is crucial in managing migraines by analyzing patient-reported symptoms, health records, genetic markers, and environmental triggers. Advanced techniques like machine learning and predictive modeling can predict attacks, optimize treatment plans, and improve patient outcomes, ultimately enhancing migraine care quality.

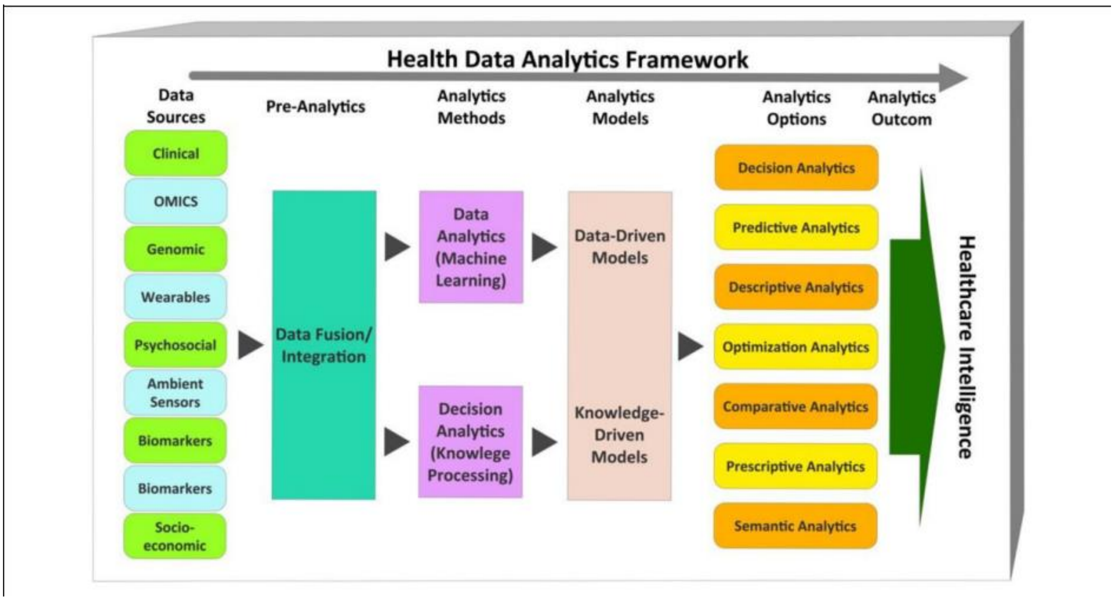


Figure 13: Health data analysis framework.

3.1 Nature of Healthcare Data

Medical data can be collected through many heterogeneous sources to create a comprehensive understanding of a patient. Some of these sources are as follows: [26]

• Biometric data which includes X-rays, CT-scan, MRIs, Sonography among other medical imaging techniques, fingerprints, retinal scans, various signals like Blood Pressure and ECG, DNA, and genetic data, etc.

• Data generated by Medical staff like Electronic Medical Records (EMR), Doctor’s notes and other paper documents

• Wearable fitness devices that monitor overall fitness levels by tracking things like steps taken, sleep time, heart activity, and calories spent

• Data captured from various sources on the internet like social media, medical websites, and blogs amongst others Data from Pharmacies.

There are plenty of opportunities for Big Data technologies to be used in the healthcare space.

By analyzing historical medical patterns of patients across various diseases and various approaches used by doctors for treatment, it is possible to make informed decisions from a broader point of view rather than an isolated one.

Public healthcare budgets can be reduced by identifying where the resources are best needed for preventive healthcare. This can be achieved by identifying which strata of society are affected the most by certain types of diseases and developing actionable insights on where prevention and education are needed the most to produce robust populations. Remote and virtual healthcare can be promoted using Big Data hence impacting the lives of millions of people without access to quality healthcare. Using clustering techniques, identification, and segmenting of patients with similar symptoms, indications, and care patterns over longer periods can help form patient clusters resulting in patient-centric medical approaches.

3.2. Examples of Data Analysis for Healthcare:

Example	Description
Asthma polis	For the treatment of asthma, the company has created a tracker called global positioning system (GPS) that monitors the usage of inhalers by patients. A small cap-like device is to be placed at the top of the inhaler that at as a sensor and helps in providing useful information. The patient that is using such type of inhaler when any time suffers from an asthma attack and uses hid inhaler that time the device will record the time and place and convey the information to the website. This data is then made available to the Center for Disease Control. The CDC takes the survey for why and from which allergic source the attack of asthma was caused to the patient. Thus, all the relevant data about the attack is gathered through the help of the device. The benefit of this device to the user is that he can generate a report of his attack and will be aware from that the source he is facing an attack of asthma.

Diabetes and big data:	In the big data revolution diabetes patients have also got up with a lot of benefits. Common Sensing company has given the GoCap application, which not only helps in recording the daily dosage of insulin but also at what time the dosage is given to the patient. This information is then fed over mobile devices where patients and other members can get this information. Thus, data becomes easier for other healthcare professionals to access and allows them to identify where the problem is and what can be the proper diagnosis for it. Another technology that has emerged with the combination of diabetes and Big Data is served by Allazo Health. Predictive analytics is being used to improve the medication program in this system.
Battling the flu	Center for Disease Control has become a strong pillar in big data for influenza. Over week 6, 80,000 flu reports are received by the CDC. All these reports which CDC gathers include the reason for the patient's sickness, what treatments are given to them, and whether that treatment is effective or not. The CDC helps the general public to make this information available to them. Doctors also get the benefits of this by getting a clearer picture of how and why the disease is spreading across the world. It helps the caretakers get information about vaccines and other antiviral medicines that can be given to patients for faster recovery. This application of big data is not only restricted to doctors use only but the patient can himself assists for better recovery. FluNearYou an application made by the Skoll Global Threats Fund and the American Public Health Association motivates users to input their symptoms before they fall sick completely thus proper diagnosis is given at a much earlier stage.

Table 03: Example of big data analysis. [26]

Conclusion

In this chapter, we have presented the notions and terms used in data analysis and the different models of machine learning and data analysis to facilitate the understanding of this work. We also highlighted the uses of data analysis in healthcare, the nature of healthcare data, and applications and examples of big data in healthcare. Among the applications of data analysis in the field of health is the classifications of migraines, which we will present in the next chapter, in which we cite the different works carried out by researchers to predict the types of migraine.

Chapter II:
State-of-the-Art

Introduction

This chapter explores the art of migraine-type prediction using our clever and intuitive sets, which have intuitive links between traditional techniques and knowledgeable professionals in this field. Predicting the types of migraines is important for managing the consequences of medication for this medical condition. New avenues for the development of predictable models are provided by recent developments in automatic application techniques.

1. Migraine

Migraine is a common primary headache disorder characterized by attacks of debilitating headaches. Although its pathophysiology remains poorly understood, there is a suggested dysfunction of the brain in regulating pain and external stimuli, as in other chronic pain syndromes. Around one-third of migraine headaches are preceded by a visual, auditory, or somatosensory aura and a majority are associated with nausea, vomiting, and sensitivity to light. Currently, diagnostic pathological changes have not been identified, and diagnosis relies on retrospective patient reports of headache characteristics. However, the heterogeneity in individual patient symptoms, along with a long list of differential diagnoses may sometimes complicate accurate diagnosis, accounting for ongoing underdiagnosis and undertreatment of this chronic condition. Further research is necessary to elucidate the pathophysiology of migraine and develop biomarkers for diagnosis.[27]

1.1 Migraine Triggers

Migraine triggers are factors or events that can precipitate or exacerbate migraine attacks in susceptible individuals. Triggers vary widely among individuals, and not all migraine sufferers are affected by the same triggers. Here are some common migraine triggers:

A. **Dietary Triggers in Migraine:** Several studies have explored the relationship between migraine and food allergies and have proposed treating migraine with specific diets such as elimination Fasting and caffeine withdrawal are two of the most common migraine triggers, but several studies have evaluated the role of diets and particular foods or additives as triggers of migraine. Common dietary triggers include Caffeine, Alcohol, particularly red wine, Aged cheeses, processed foods containing additives like MSG (monosodium glutamate) or nitrates/nitrites, Chocolate, and Citrus fruits. [28]

B. **Menstrual Changes as a Migraine Trigger:** Diary studies confirm that menstruation can trigger migraine, perhaps more so in those with aura. Migraine attacks during menses may be more severe, with reduced response to acute medication such as triptans. Estrogen withdrawal before menses is likely the cause of menstrual migraine and explains why migraine onset often commences before menstruation onset or on the first day. For a significant minority of women, menstrual-related migraine may be more likely to occur towards the end of their cycle suggesting a relationship to blood loss or anemia. [28]

C. **Weather Triggers in Migraine:** Multiple studies have considered the relationship between migraine or headache and environmental factors such as weather. High altitude is a proven trigger of headaches especially during rapid ascents. Living at high altitudes may cause chronic headaches, especially in those lacking genetic adaptations to adapt to hypoxia. [28]

D. **Sensory Stimuli as a Migraine Trigger:** Visual, noise, olfactory, and other sensory stimuli exacerbate migraine intensity and may cause migraine in vulnerable individuals. Individuals with migraine have lower thresholds of discomfort to stimuli such as lights, sounds, odors, and thermal and mechanical stimulations.

E. **Stress as a Migraine Trigger:** Stress is perhaps the most common self-reported trigger of migraine, and many studies have been able to demonstrate a link between chronic stress, pain, migraine, and catastrophic thinking. [28]

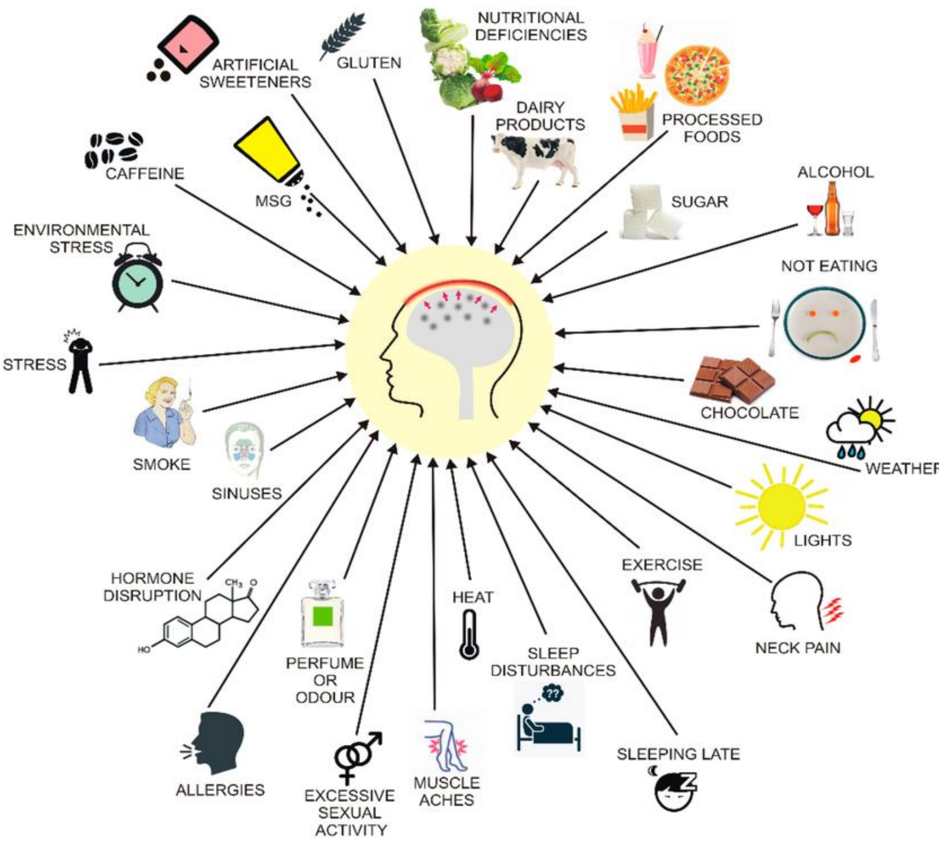


Figure 14: migraine triggers.

1.1 Migraine Types

Before we discuss the types of migraines, it's important to acknowledge the diverse nature of this neurological disorder. Migraines present in various forms, each with unique symptoms and characteristics. Let's explore these different types to gain a better understanding of the condition.

- **Hemiplegic migraine**

A hemiplegic migraine is a rare form of migraine in which the migraine headache attack is accompanied by unilateral weakness. Typically, migraine aura has visual symptoms, but motor symptoms are rare. Hemiplegic migraine is a rare subtype of migraine with aura, characterized by the presence of motor weakness as aura manifestation. The motor weakness is often accompanied by other forms of aura, like impairment in vision, speech, or sensation. Hemiplegic migraine can run in the family (familial hemiplegic migraine) or occur sporadically in one individual (sporadic hemiplegic migraine). [28]

- **Migraine with aura (also called classic migraine):**

Is a recurring headache that strikes after or at the same time as sensory disturbances called aura. These disturbances can include flashes of light, blind spots, and other vision changes or tingling in your hand or face. [29]

- **Familial hemiplegic migraine (FHM):**

Is an autosomal dominantly inherited subtype of migraine with aura, characterized by transient neurological signs and symptoms. Typical hemiplegic migraine attacks start in the first or second decade of life. Some patients with FHM suffer from daily recurrent attacks since childhood. [30]

- **A basilar migraine**

Basilar migraine is rare. Because it begins in the brain stem, a doctor may call it “migraine with brain stem aura.” This aura may involve changes in a person’s speech, hearing, or vision they may see lines, flashes of light, or spots in their field of vision. Also, a person may experience pain before or during these changes, and the pain may occur on one or both sides of the head. [31]

- **Silent migraine**

With this kind, also known as an acephalgic migraine, you have aura symptoms without a headache. You may also have nausea and other migraine symptoms. An attack usually lasts only about 20-30 minutes. [32]

- **Sporadic hemiplegic migraine (SHM)**

  Is defined as migraine attacks associated with some degree of motor weakness/hemiparesis during the aura phase and where no first-degree relative (parent, sibling, or child) has identical attacks. [33]

2. Role of machine learning in migraine prediction:

A valuable strategy to reduce the unmet needs in the understanding of primary headaches is to study headache patients using machine learning approaches. These methods have been employed to study patients with neurological or psychiatric conditions—like Alzheimer's disease, depression, and chronic pain disorders—to identify neuroimaging biomarkers, which could be used to predict clinical outcomes, including diagnostic categories, measures of symptoms, prediction of disease evolution, and treatment response. There are two main machine learning approaches: supervised and unsupervised. Supervised machine learning algorithms are trained to automatically classify individuals into predefined groups, e.g., patients or healthy controls, and yield an associated accuracy indicative of how well the model could generalize to future individual cases. At a more detailed level, a machine learning classifier is a function that takes the values of various features (e.g., different imaging patterns) in an example and predicts the class that the example belongs to (e.g., patient or control). The goal is to develop a "classifier" that identifies the relation between each example and its respective category with high accuracy. Based on what the algorithm has learned, it will be then able to classify new, previously unseen data into one of the predefined categories. By contrast, unsupervised machine learning models are data-driven automated approaches that, without the availability of a priori information supplied by the operator, seek to classify uncategorized data, with the primary aim of discovering unknown, but potentially useful information in the data. These classification models include a "training" phase in which training data are used to develop an algorithm able to discriminate between groups and a "testing" phase in which the algorithm is used to blind-predict the group to which a new observation belongs. The main advantages of using machine learning approaches are that they allow inference on an individual patient basis and are sensitive to subtle and spatially distributed patterns of disease-induced changes in the brain that might be undetectable at group-level comparisons. The evaluation of the performance of the model in a new subset of individuals provides valid estimates of how well the discriminative model generalizes to new data, enhancing the clinical significance of these approach metrics that disclose complementary information regarding the underlying biological processes. [34]

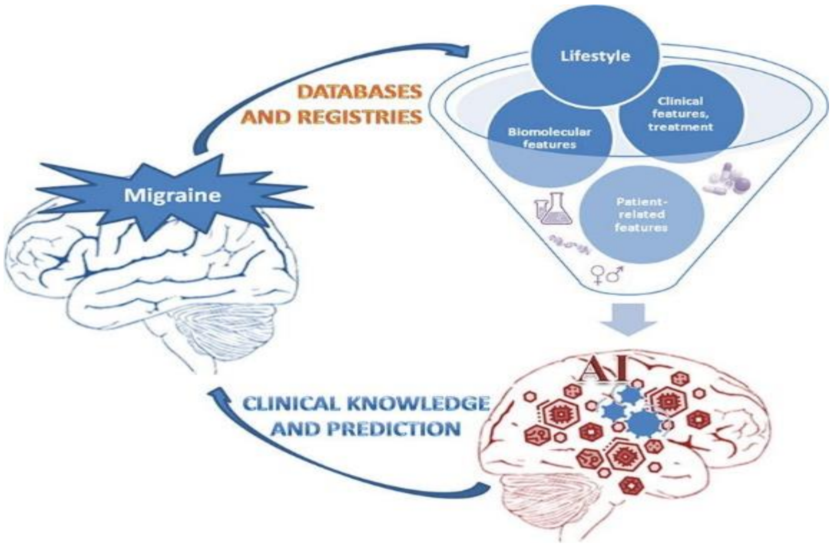


Figure 15: migraine prediction.

3. Migraine prediction

The unpredictable nature of migraine attacks creates significant challenges for individuals living with episodic migraine. In theory, accurate real-time forecasts of oncoming migraine attacks could help patients target medication use and other preventive strategies for susceptible periods and help decrease the disease burden. Several different types of data may correlate with day-to-day migraine risk, and thus may help to predict migraine; these include trigger factors, premonitory symptoms, self-prediction, and physiological signals. Though trigger factors have been formally studied for at least 60 years, and certain transient exposures (e.g., psychological stress, changes in sleep, and menstrual cycles) are commonly thought to bring about migraine attacks, as of yet there is limited work toward assessing their predictive utility. [35]

5. Migraine prediction methods

Migraine prediction involves identifying triggers, tracking symptoms, and observing patterns. Techniques include keeping a migraine diary, recording triggers like dietary factors and environmental influences, and recognizing patterns in the timing and occurrence of attacks. This helps in predicting when an attack is likely to occur and taking preventive measures. Stress management, sleep hygiene, and medication usage are also crucial components of classic migraine prediction methods.

4.1 Classic methods for migraine prediction:

Classic methods of predicting migraines often involve tracking triggers, symptoms, and patterns. Here are some common approaches:

A) Trigger Identification:

Record triggers like certain foods, beverages, environmental factors, stress levels, hormonal changes, and sleep patterns to prevent migraines by identifying and avoiding them.

B) Tracking Symptoms:

Maintain a migraine diary to document the onset, duration, severity, and symptoms of each attack, including visual disturbances, nausea, vomiting, sensitivity to light or sound, and mood changes.

C) Pattern Recognition: Identify patterns in migraine attacks, such as specific times of day, week, or menstruation, to predict their likelihood and manage migraines effectively.

D) Medication Tracking: Migraine medication should be taken at the earliest sign of a migraine attack, as triptans are more effective when taken at this time, allowing for more accurate timing of medication use.

Many traditional methods for predicting migraines have drawbacks due to their simplicity, subjectivity, and inability to predict migraine occurrences accurately. These methods rely on basic algorithms and heuristic rules, which may not be enough due to the multifactorial nature of migraines and individual variability. They also struggle to capture the complex interactions among physiological, environmental, and genetic factors that influence migraines, resulting in limited accuracy and generalizability across diverse patient populations. Moreover, they struggle to adapt to new data sources or real-time information, which makes them less scalable and usable in clinical practice**.**

6. Related works and comparative study

In this section, we reference related work that employs machine learning techniques for classifying migraine types and their comparative study.

6.1. Related works

The study in [36], aims to improve the detection and classification of migraine phenotypes with aura (MwA) by using advanced neuroimaging techniques and machine learning algorithms. Participants with MwA and healthy controls are involved, and various features such as cortical thickness, surface area, volume, mean Gaussian curvature, and folding index are extracted. The study creates two datasets for machine learning algorithm training, with features selected using the Extremely Randomized Trees algorithm. The study applies various machine learning algorithms, with hyperparameter tuning for optimization. The study involves 78 subjects, and the results demonstrate promising performance in classifying participants based on their health status and migraine category. The objectives include developing accurate machine learning algorithms for MwA detection and subtype classification, investigating cortical abnormalities and biomarkers associated with MwA pathophysiology, evaluating algorithm performance, validating accuracy and reliability, and exploring potential treatment targets. [36]

This research in [37]: seeks to: develop a comprehensive methodology for classifying migraines into two categories: aura-positive and aura-negative migraines. The data is extracted from a migraine-related patient database, and five classification models, namely multilayer perceptron-type artificial neural network (MLP), logistic regression, support vector machine (SVM), nearest neighbor, and decision tree are used for pre-processing the data. The model achieves precision levels of 97% with all variables and 98% with a reduced set of 18 variables. The study concludes that artificial neural networks are powerful tools for migraine classification, offering high precision and accuracy. Future research may explore different neural network architectures and analyze the relationship between clinical diagnosis and theoretical classification of migraine types. [37]

This project in [38] intends to: delve into the creation and validation of an algorithm that classifies primary headache disorders, such as migraines and tension-type headaches (TTH), utilizing the International Classification of Headache Disorders (ICHD-3) criteria. The algorithm is specifically developed to categorize individual headache events in a migraine management application's database, with its performance verified against classifications made by a specialist. The algorithm is designed to account for differences in pain intensity scales and differentiate between probable and definite diagnoses. The study involved collecting data from 102 headache events, with any cases of disagreement carefully analyzed for improvement. While the study highlights the algorithm's potential as a diagnostic tool, it also acknowledges limitations in data collection methods and potential biases. Future research could integrate electronic headache diaries into clinical practice, use real-world databases for epidemiological research, and refine the algorithm to incorporate additional diagnostic functionalities. [38]

This work in [39] is designed to: Digital health technologies like telehealth, wearables, and apps are changing the way healthcare is delivered by providing solutions for chronic conditions and improving access to care. Among these technologies, digital tools have proven to be beneficial for headache disorders such as migraines, as they reduce diagnostic delays and enhance management strategies. These tools promote patient engagement and shared decision-making, thereby transforming the delivery of healthcare.

In a systematic review of 41 studies, a considerable increase in computerized migraine diagnostic tools over the last decade was observed, with high median diagnostic accuracy values. However, most evaluations were restricted to headache clinical centers, which limits their generalizability to community or primary care settings. The review acknowledges limitations such as tool heterogeneity, potential incorporation bias, and language restrictions. Despite these limitations, digital health technologies, particularly computerized migraine diagnostic tools, hold promise in transforming headache care delivery and improving patient outcomes. [39]

The study in [40]: A project was undertaken to train machine learning models for predicting migraine occurrences using data from an online public database. To reduce bias, the Synthetic Minority Over-Sampling Technique (SMOTE) was employed. Two models were trained, namely logistic regression and random forest. The logistic regression model performed with 97% accuracy, while the random forest model achieved 89%. The study identified triggers and helping factors such as sleep or exercise as crucial factors. As a result, a user-friendly website was developed to display migraine probability using logistic regression. [40]

The research in [41] attempts to: assess the accuracy of immunos algorithms in diagnosing migraine, tension-type, and cluster-type headaches in Turkey. The research involved evaluating 850 patients via an expert system based on a website survey. The results of the study indicated a maximum accuracy of 95.65%, demonstrating the potential of immunos algorithms to classify primary headaches accurately. The system's ease of information sharing and precise results will be beneficial for neurologists. [41]

[42]: This study tested feature selection and machine learning classification methods to automate migraine diagnosis using DTIs and emotion and cognition-related questionnaire answers. Fifty-two adult subjects, divided into control, sporadic migraine, and chronic migraine with medication overuse groups, underwent magnetic resonance imaging to assess white matter integrity. The DTI images and test results were processed using feature selection algorithms (Gradient Tree Boosting, L1-based, Random Forest, Univariate) and classification algorithms (SVM, Boosting, Naive Bayes). A committee method was implemented to enhance classification accuracy. This approach significantly improved classification accuracy, achieving over 90% accuracy across all classifiers. Key features included those related to pain, analgesics, and the left uncinate brain. The findings suggest that the proposed committee method effectively supports specialists in diagnosing migraines using MRI data.

[43] This prospective, observational study aimed to assess patients' ability to predict migraine attacks based on their PSs and to identify any characteristic profiles of good predictors. Patients with migraine, with and without aura, recorded their perceived PSs and subsequent pain onset using a mobile app over two months. Out of the initial 50 patients, 34 were analyzed, documenting 229 attacks. PSs were recorded in 69% of attacks, with 27.5% documented before pain onset. While 67.6% of patients predicted at least one attack, only 35.3% were good predictors (accurately predicting >50% of attacks). The positive predictive value was 85.1%, with only 11 false reports. Good predictors did not share specific clinical characteristics, though symptoms like photophobia, drowsiness, yawning, increased thirst, and blurred vision were particularly predictive. Most patients experienced PSs and could predict at least one attack, although only a few were consistently accurate without any distinguishing profile.

[44] The study involved a customized ML-based decision support system combining support vector machines and Random Optimization (RO-MO). Using a dataset of 777 consecutive migraine patients, we identified predictors with higher discriminatory power for MO than baseline SVM. The top four predictors were integrated into the final RO-MO system, stratifying risk into five levels. The ROC analysis yielded a c-statistic of 0.83, with sensitivity and specificity of 0.69 and 0.87, respectively, and an accuracy of 0.87 when MO was predicted by at least three RO-MO models. Logistic regression confirmed the effectiveness of the RO-MO system, with odds ratios of 5.7 and 21.0 for patients classified as probably or definitely at risk of MO, respectively. In conclusion, combining ML and RO, considering clinical and biochemical features, drug exposure, and lifestyle, represents a promising approach to MO prediction in migraine, enhancing model precision by weighting attribute importance.

[45] This study aimed to analyze each EEG channel separately in migraine patients, focusing on the increase in magnitude under flash stimulation. The beta band data from each EEG channel were pre-processed using the Burg-AR method and then classified with a support vector machine (SVM) to determine the most definitive channels. The study found that the T3, F7, O1, and O2 channels were the most decisive for migraine diagnosis based on PSD magnitude increases under flash stimulation. Additionally, it identified which brain lobes are more affected by migraine triggers and confirmed the asymmetry feature of migraine through EEG. The study suggests alternative migraine diagnosis methods for future research, based on the physiological response of the scalp to flash stimulation.

[46] This thesis uses machine learning to improve migraine diagnosis by using binary and multiclass classification techniques. Data from a previous study is preprocessed using models like Random Forest, Logistic Regression, and SVM. The goal is to minimize false negatives and improve sensitivity, specificity, and accuracy. Results show that machine learning techniques significantly improve migraine diagnosis.

[47] The study introduces a high-performance Headache Prediction Support System (HPSS) for preliminary diagnostic guidance. The system uses a hybrid machine learning model and combines clustering and classification to improve diagnostic results. The system achieved the highest accuracy of 99.1% for migraine prediction and 93% overall. A web-based interface is available for easy use. The study also evaluates the impact of headache symptoms on diagnosis, aiding medical experts in refining criteria.

[48] The study aims to create a predictive model for migraine symptoms, enabling early management. It evaluates the accuracy of machine learning algorithms in predicting migraines and their effectiveness in predicting attacks. The model uses machine learning software to analyze temperature, heart rate, and EEG signals, identifying migraines. The findings suggest the model can accurately diagnose migraines.

6.2. Comparative study

Article	Author	Method	model	Data Set used	Accuracy
[42]	Yolanda Garcia-Chimeno	Machine Learning classification	SVM Boosting (AdaBoost) Naive Bayes	Private Data Set	Naïve Bayes 67 to 93% Boosting 93 to 94% SVM 90 to 95%
[43]	Ana B Gago-Veiga	Machine Learning classification	Application (BrainGuard App)	Rami L, Valls-Pedret C, Bartrés-Faz D, et al.e. Rev Neurol. 2011;52(4):195–201	85,1%
[44]	Patrizia Ferroni	Machine Learning classification	SVM Random Optimization (RO-MO)	Private data set	SVM 81% RO-MO 87%
[45]	Selahaddi Batuhan Akben Ahmet Alkan- Deniz Tuncel-	Machine Learning classification	SVM	data is obtained from Neurology Department of Kahramanmaras Sutcu Imam University	88.4%
[46]	Burak Ozdemir Supervisors: Dr. M. van Leeuwen & MSc I. Papagianni	Machine Learning	Random Forest, Logistic Regression, and SVM	LUMINA questionnaire	Random forest 89% Logistic regression 89% SVM 79%

[47]	Ahmad Qawasmeh	hybrid machine learning algorithm.	Employ a hybrid machine learning algorithm that consists of clustering and classification to train a classifier, while using 26 different classification .algorithms	collect dataset by asking real patients to fill out a headache questionnaire based on what they feel.	K-Means and Random Forest with a migraine accuracy of 99.1% and an overall accuracy of 93%
[48]	Hye-Kyeong Ko	Machine Learning	CNN Classification models	Private (the data were standardized and combined to show trends and relationships between the subjects' EEGs and EKGs and the frequency of migraine attacks)	99,06%

Table 04: Comparative study of related works.

The previous table provides a comprehensive summary of various studies and applications aimed at predicting and classifying migraine episodes. Researchers have used different datasets, techniques, and applications to accomplish this task. They have used a range of methodologies, including the analysis of private datasets that encompass factors such as emotion, cognition, and pain perception, as well as the use of diffusion tensor images (DTIs) and questionnaire responses. The techniques used have varied from traditional feature selection algorithms such as Gradient Tree Boosting and Random Forest to more advanced classification models such as Support Vector Machines (SVM), AdaBoost, Naive Bayes, and Convolutional Neural Networks (CNN). The applications developed have ranged from mobile apps designed to track migraine symptoms and identify potential triggers to machine learning-based decision support tools that aid healthcare professionals in diagnosing and managing migraine patients. Overall, these studies highlight the multidimensional approach taken by researchers to address the complexity of migraine prediction and classification, with the ultimate goal of improving patient care and treatment outcomes.

7. Synthesis

- This study revealed that migraine types are frequently predicted by machine learning algorithms, particularly those that are based on supervised learning.

- Using artificial intelligence to discover complex patterns in data to determine the distinctive characteristics of each type of migraine.
- Contribute to making results interpretable so that doctors and patients can understand expectations and act accordingly.
- Make interpretations of predictions transparent and understandable to ensure confidence in clinical decisions.
- Enabling a better understanding of the condition and more targeted and effective interventions for patients.
- In the referenced works, researchers concentrate exclusively on predictive analysis. However, we can also leverage other types of analysis, such as descriptive and diagnostic analysis, to extract valuable insights from raw data and comprehend the connections between various symptoms and types of migraines.

Conclusion

In conclusion, this chapter provided an overview of the state of the art of predicting migraine types, focusing on the datasets used, related works, and classical and intelligent methods employed. Recent advances in machine learning techniques have opened new avenues for the accurate prediction of types. However, it is important to consider the specific characteristics and limitations of each method, as well as the availability of high-quality data. This literature review will allow us to better understand existing approaches and lay the foundations for our research work on predicting migraine types, combining the advantages of classic and intelligent methods to obtain precise and reliable results.

Chapter III: Proposed Approach and Results

Introduction:

This chapter will focus on developing and implementing our migraine-type prediction system. We will discuss various technical aspects of our project, including the hardware, development tools, choice of algorithm and operation, and the selected dataset. Additionally, we will describe the architecture of the modeling steps for the prediction and discuss the results obtained, comparing them to similar works.

1. Development tools

My development tools encompass both hardware and software components.

1.1 Hardware Tools:

The development of the model is carried out via laptop with the following characteristics:

Mark	Dell Latitude E5470
Processor	Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz 2.50 GHz
RAM	8,00 GB
Hard disk	500 GO
Operating system	Windows 10 Pro

1.2 Software tools

In terms of software tools, I rely on a diverse range of applications and platforms tailored to specific tasks within the realm of data analysis, machine learning, and software development. This includes:

- **Python:** Python is a high-level programming language with dynamic semantics, making it ideal for Rapid Application Development and scripting. Its simple syntax emphasizes readability, reducing maintenance costs. Python supports modules and packages, encouraging modularity and code reuse. The interpreter and standard library are available in source or binary form for all major platforms. Python's fast edit-test-debug cycle makes debugging easy, with a source-level debugger allowing inspection of variables, evaluation of expressions, and code stepping. The fast edit-test-debug cycle makes it an effective debugging tool.[49]

- **Jupyter Notebooks:** Jupyter Notebooks provide an interactive computing environment ideal for data exploration, visualization, and experimentation. They allow for the creation of documents that combine live code, equations, visualizations, and narrative text.
- **Integrated Development Environments (IDEs):** I often utilize IDEs such as PyCharm, Visual Studio Code, or JupyterLab for writing, debugging, and testing code efficiently.
- **Various Data Visualization Libraries:** Depending on the requirements of a project, I leverage libraries like Matplotlib, Seaborn, Plotly, or Bokeh for creating insightful visualizations to communicate findings effectively: [50]

  Matplotlib: This open-source library in Python is widely used for publishing quality figures in various hard copy formats and interactive environments across platforms. You can design charts, graphs, pie charts, scatterplots, histograms, error charts, etc., with just a few lines of code.

  Pandas: It is an open-source, BSD-licensed library. Pandas enable the provision of easy data structure and quicker data analysis for Python. For operations like data analysis and modeling, Pandas makes it possible to carry these out without needing to switch to a more domain-specific language like R.

  Numpy: is one of the fundamental packages for Python, providing support for large multidimensional arrays and matrices along with a collection of high-level mathematical functions to execute these functions swiftly. NumPy relies on BLAS and LAPACK for efficient linear algebra computations. NumPy can also be used as an efficient multi-dimensional container of generic data.

  Keras: It is an open-source neural network library written in Python designed to enable fast experimentation with deep neural networks. With deep learning becoming ubiquitous, Keras becomes the ideal choice as it is API designed for humans and not machines, according to the creators.

  Seaborn: When it comes to the visualization of statistical models like heat maps, Seaborn is among the reliable sources. This Python library is derived from Matplotlib and is closely integrated with Pandas data structures. Visit the installation page to see how this package can be installed.

  Tensorflow: TensorFlow's most popular deep learning framework is an open-source software library for high-performance numerical computation. It is an iconic math library used for Python in machine learning and deep learning algorithms. TensorFlow was developed by the researchers at the Google Brain team within the Google AI organization. Today, it is being used by researchers for machine learning algorithms and by physicists for complex mathematical computations.

These tools collectively facilitate the development, analysis, and deployment of machine learning models and other software applications, enabling efficient and effective workflows throughout the development lifecycle.

2. Proposed approach

AI could support migraine diagnosis and management in many ways. For instance, it could help non-headache specialists reach the correct diagnosis and guide them to the choice of the best treatment for the patient [36]. In this work, we propose a data-driven approach based on data analysis and machine learning for migraine classification. The proposed approach is summarized in Figure 15.

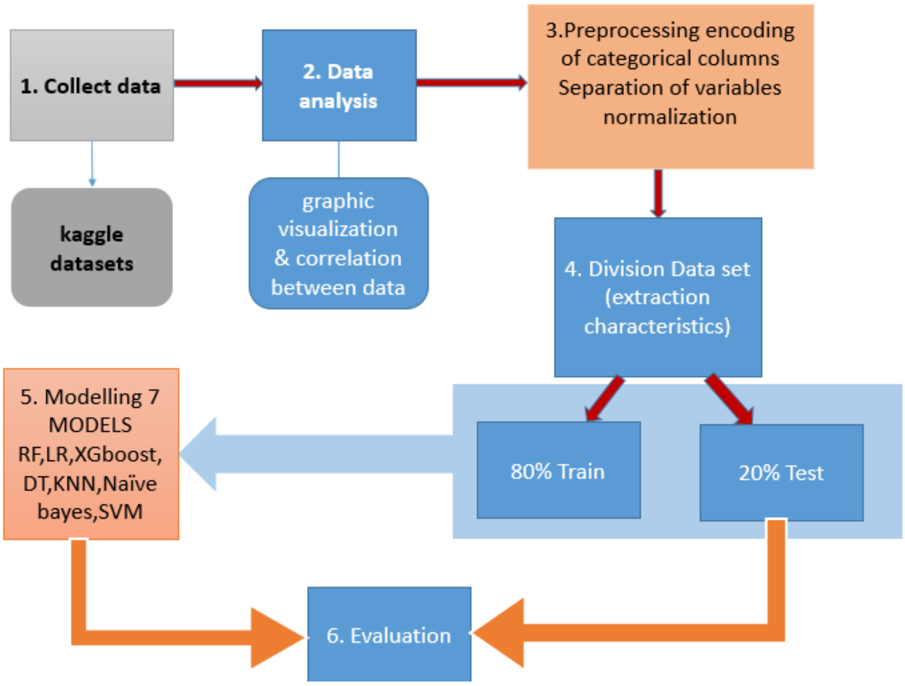


Figure 16: Architecture of the prediction steps of migraine type.

Our approach mainly consists of 6 phases:

2.1 Data Collection

In this phase, our objective is to identify relevant information that can aid in resolving the issue or enhance the analytical approach to achieve our goal of predicting migraine types. This step involves examining available data sources and pinpointing pertinent data, which may require various tasks such as querying databases, extracting information from data feeds, submitting requests to other departments, or searching for external data sources. The forms of data resources can vary, with some companies allowing data analysts access to their data to analyze the company's past, present, or future actions to optimize profits.

To accurately predict migraine types, it's crucial to gather data from various sources. These sources can include clinical records detailing past diagnoses and treatments, symptom-tracking apps providing real-time insights into headache frequency, duration, and triggers, wearable biometric sensors capturing physiological markers such as heart rate variability and sleep patterns, genetic testing revealing predispositions to specific migraine subtypes, environmental data on factors such as weather changes and stress levels, patient-reported outcomes via surveys, and electronic health records providing comprehensive health histories. By analyzing and integrating these diverse data sets, experts can develop predictive models to anticipate migraine types and improve personalized treatment strategies.

The data for our study was sourced from Kaggle datasets, and a snapshot of the Excel file is provided below: data migraine.csv [51]

The dataset comprises 400 medical records of users diagnosed with various pathologies associated with migraines. Data were recorded by trained medical personnel at the Centro Materno Infantil de Soledad during the first quarter of 2013. The compiled database contains information regarding symptoms or variables of interest required for the classification of migraines.

2.2. Descriptive and diagnostic data analysis

At this stage, to obtain insights, reach conclusions, and detect any issues related to the data, an analysis and visualization phase is recommended. A descriptive analysis technique is proposed for the discovery of data. Various data analysis techniques are available to understand data, as detailed in the first chapter. Data visualization can also be utilized to analyze data in the form of graphs, to extract additional information that the data may contain. Data visualization refers to the graphical representation of data to make it more interpretable. These analysis techniques comprise understanding the dataset (the columns, the types of attributes, the number of samples, etc.), applying statistical methods (calculating the mean, the median, etc.), and exploratory analysis where we can view the data in different ways:

The figure 17 resumes the collected dataset, and table 05 includes the features description.

```
RangeIndex: 400 entries, 0 to 399
Data columns (total 24 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Age          400 non-null    int64
 1   Duration     400 non-null    int64
 2   Frequency    400 non-null    int64
 3   Location     400 non-null    int64
 4   Character    400 non-null    int64
 5   Intensity    400 non-null    int64
 6   Nausea       400 non-null    int64
 7   Vomit        400 non-null    int64
 8   Phonophobia  400 non-null    int64
 9   Photophobia  400 non-null    int64
 10  Visual       400 non-null    int64
 11  Sensory      400 non-null    int64
 12  Dysphasia    400 non-null    int64
 13  Dysarthria   400 non-null    int64
 14  Vertigo      400 non-null    int64
 15  Tinnitus     400 non-null    int64
 16  Hypoacusis   400 non-null    int64
 17  Diplopia     400 non-null    int64
 18  Defect       400 non-null    int64
 19  Ataxia       400 non-null    int64
 20  Conscience   400 non-null    int64
 21  Paresthesia  400 non-null    int64
 22  DPF          400 non-null    int64
 23  Type         400 non-null    object
```

Figure 17: Collected dataset resume.

Data field	Description	Values
Age	Patient's age	[16...70] years
Duration	duration of symptoms in last episode in days	[1,2,3] days
Frequency	Frequency of episodes per month	[1…8]
Location	Unilateral or bilateral pain location	[0,1,2]
character	Throbbing or constant pain	[0,1,2]
Intensity	Pain intensity (mild, medium, or severe)	[0,1,2,4]
Nausea	Nauseous feeling	[0,1]
Vomit	Vomiting	[0,1]
phonophobia	Noise sensitivity	[0,1]
photophobia	Light sensitivity	[0,1]
Visual	Number of reversible visual symptoms	[0…4]
Sensory	Number of reversible sensory symptoms	[0,1,2]
Dysphasia	Lack of speech coordination	[0,1]
Dysarthria	Disarticulated sounds and words	[0,1]
Vertigo	Dizziness	[0,1]
Tinnitus	Ringing in the ears	[0,1]
Hypoacusis	Hearing loss	[0,1]
Diplopia	Double vision	[0,1]
Defect	Simultaneous frontal eye field and nasal field defect and in both eyes	[0,1]
Ataxia	Lack of muscle control	0
Conscience	Jeopardized conscience	[0,1]
Paresthesia	Simultaneous bilateral paresthesia	[0,1]
DPF	Family background	[0,1]
Type	Diagnosis of migraine type	Typical aura with migraine, Migraine without aura, Typical aura without migraine,

		Familial hemiplegic migraine, Sporadic hemiplegic migraine, Basilar-type aura, Other

Table 05: description of the dataset.

- **Univariate analysis**

Descriptive analysis is considered the simplest type of analysis. It provides in-depth information about each attribute present in the dataset using univariate analysis. Univariate analysis involves analysing a single variable from a set of data to study the distribution of variables (Histogram, Bar Chart, etc.). For data distribution, graphical representations are used. Visual graphics provide a summary or detailed description of raw data that can be interpreted by humans.

Figure 18 visualizes the distribution of the dataset's features.

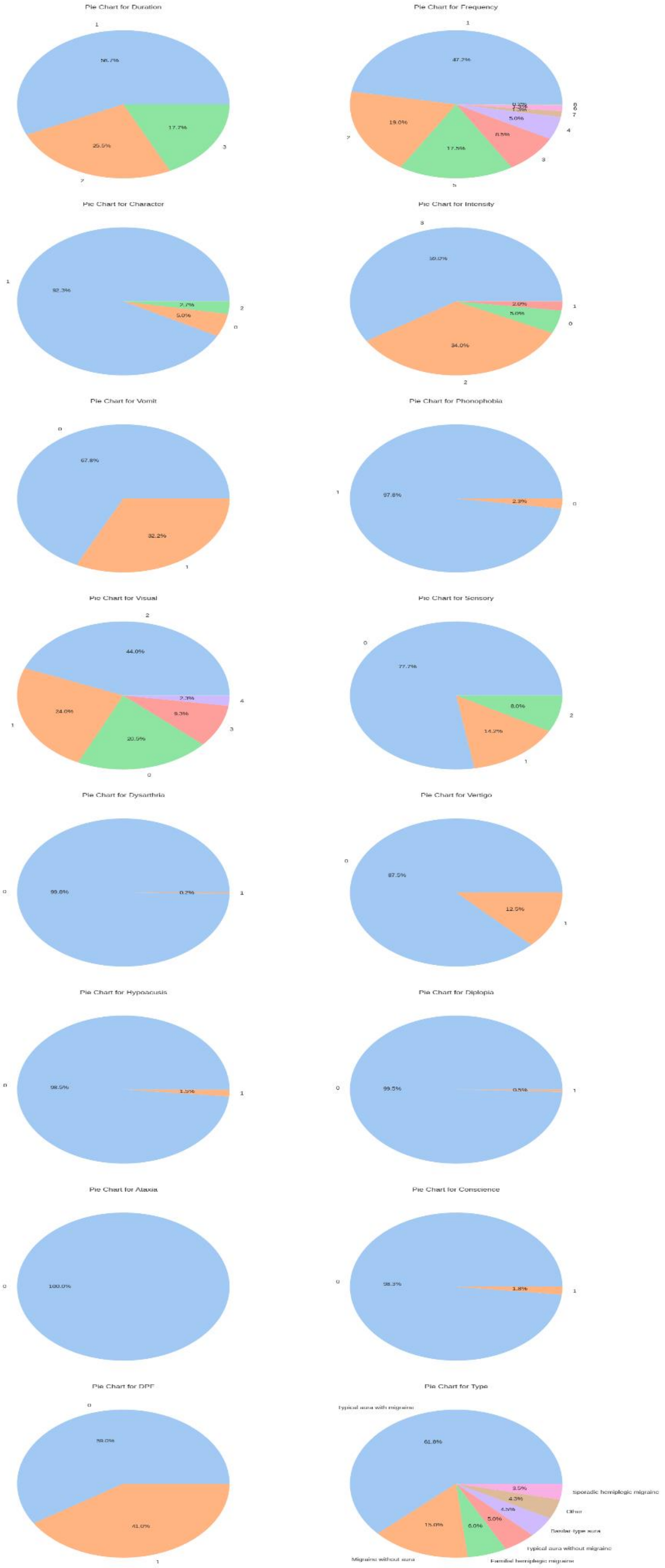


Figure 18: Data distribution.

These charts offer a clear visual representation of the prevalence and distribution of different categories within each feature. The dominant categories are easily identifiable, and the data is visualized in a way that facilitates a quick understanding of the overall distribution. These charts indicate that most of the features exhibit significant imbalance. For example:

The feature “Duration”:

- Category 1: 50.7%
- Category 2: 29.5%
- Category 3: 17.7%
- Observation: Moderately balanced, but category 1 is overrepresented compared to the others.

The figure 19, shows the individual pie chart for Type distribution:

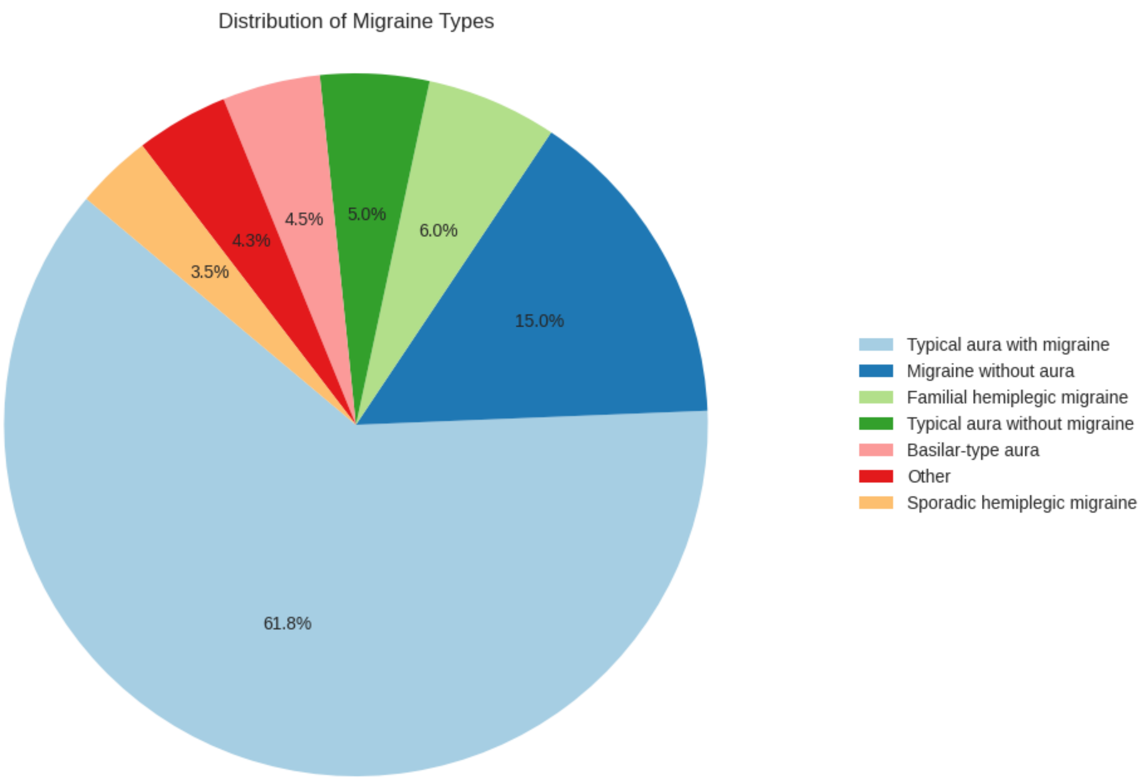


Figure 19: the individual pie chart for Type distribution.

The feature “Type”:

- Typical aura with migraine: 63.8%
- Migraine without aura: 15.0%

- Typical aura without migraine: 10.0%
- Smaller categories (Familial hemiplegic migraine, Basilar type aura, Other, Sporadic hemiplegic migraine): Lower percentages.
- Observation: Highly imbalanced, with a strong bias towards "Typical aura with migraine".

We can conclude:

- Highly Imbalanced Features: Location, Character, Nausea, Phonophobia, Photophobia, Dysphasia, Dyskinesia, Tinnitus, Hypoacusis, Defects, Ataxia, Conscience, Paresthesia. These features show a strong dominance by a single category, indicating a significant imbalance.
- Moderately Imbalanced Features: Duration, Frequency, Intensity, Vomit, Visual, Sensory, Vertigo, Diplopia, DPF, Type. These features have some degree of imbalance, but there is still representation across multiple categories.

For visualizing age distribution, we use a histogram because it has a wide range of values.

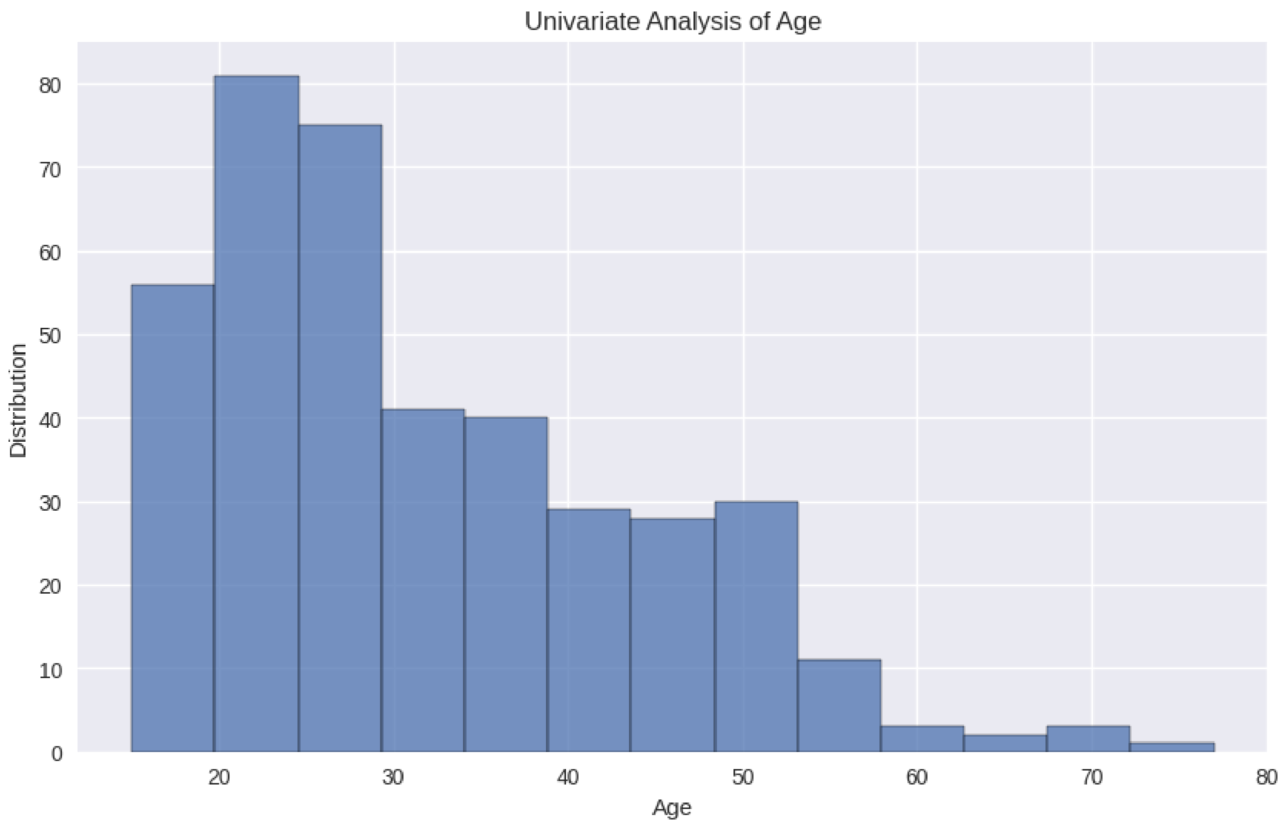


Figure 20: Age distribution.

Here are some observations based on the visualization:

The tallest bar in the histogram corresponds to the age range starting at 20. This indicates that the highest frequency of individuals occurs in their early twenties. This age group seems to be particularly important within the context of migraine occurrence or diagnosis in this specific dataset. However, there's a slight increase in frequency in the age group starting at 40. Fewer occurrences or diagnoses are observed among older age groups (e.g., 50s, 60s, 70s, and 80s) within this sample.

- **Bivariate analysis**

Bivariate analysis, on the other hand, considers two variables. The two variables can be numeric, one categorical and the other numeric.

The figure 21, illustrates the distribution of different symptoms (e.g., Photophobia, Phonophobia) across different types of migraines.

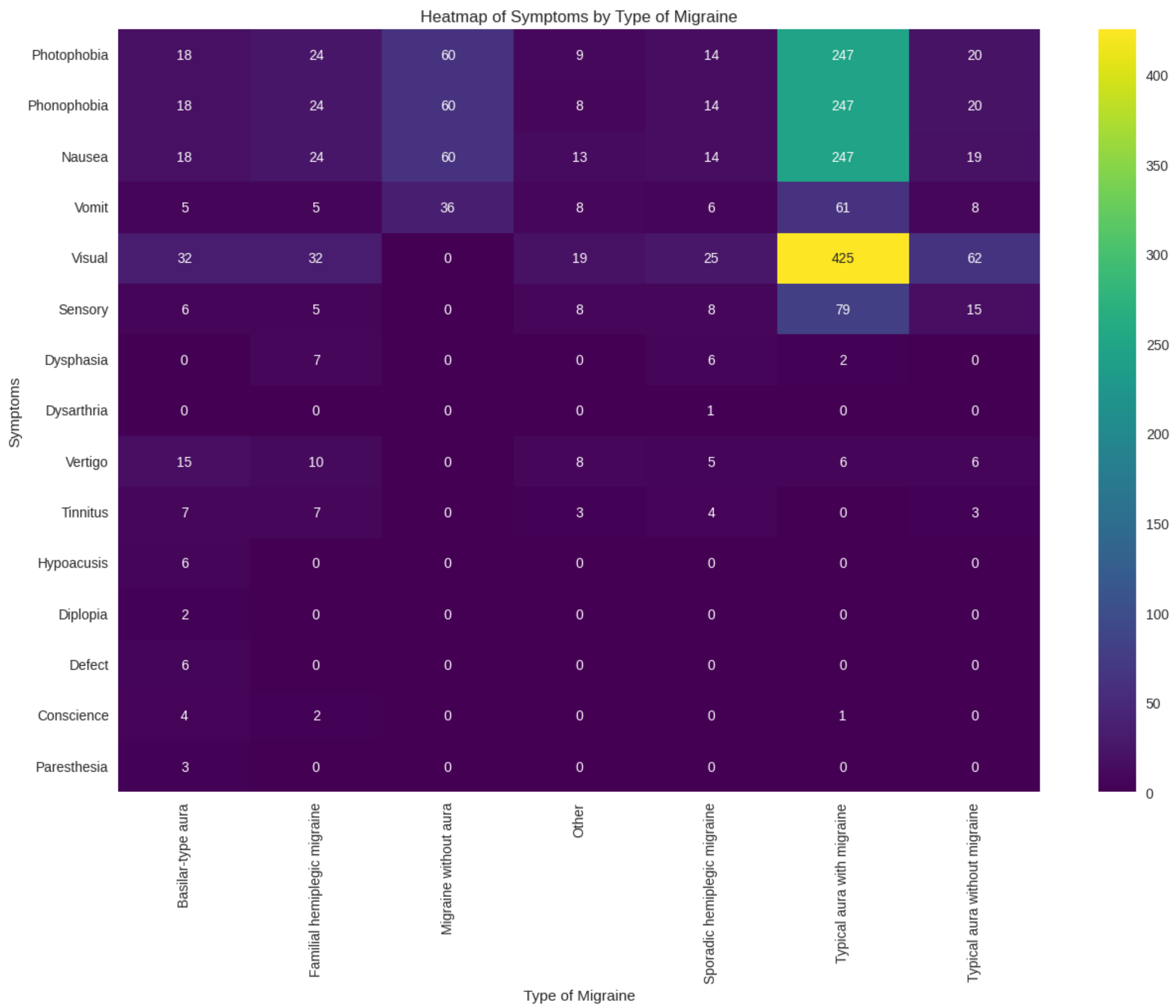


Figure 21: Heatmap of symptoms by type of migraine.

Each cell in the heatmap represents the frequency of a symptom for a specific type of migraine. Darker colors indicate higher frequencies, while lighter colors represent lower frequencies. Medical professionals can use this information to understand symptom patterns and tailor treatment strategies. Researchers studying migraine patterns can gain insights from this visual representation.

- **Multivariate analysis**

Multivariate analysis involves analyzing the data by considering more than two variables. In this case we generate a correlation heatmap for numerical features to visualize the correlation between them. The correlation heatmap is presented in figure…

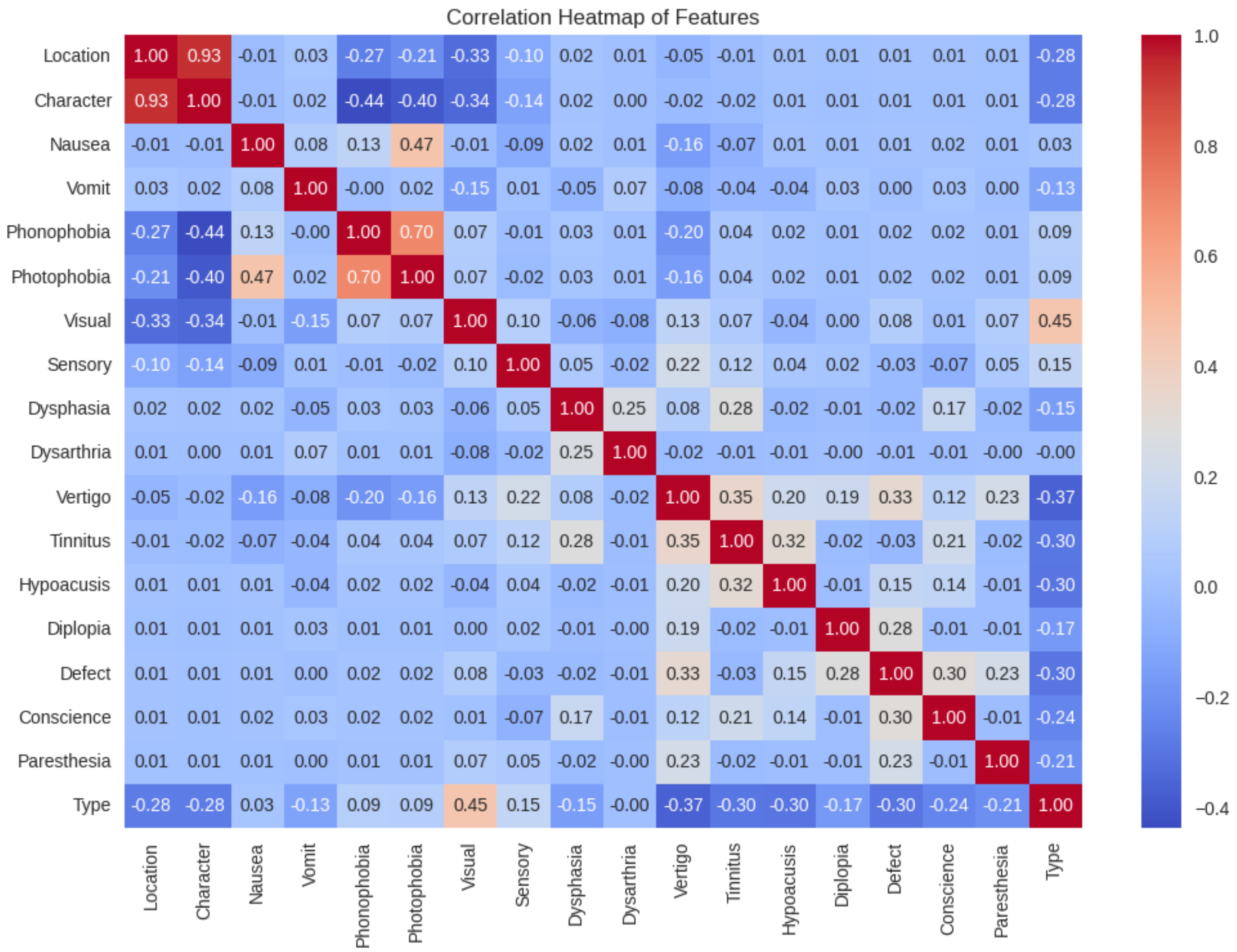


Figure 22: Correlation Heatmap of Features.

Here are some observations based on the heatmap:

- Diagonal (Self-Correlations): The diagonal from the top left to bottom right shows perfect positive correlations (value 1.0) for each feature with itself. This is expected since any feature correlates perfectly with itself.

- Positive Correlations: Features that are positively correlated (indicated by warmer colors, such as red) tend to increase or decrease together. For example, if one attribute increases, the other tends to increase as well.
- Negative Correlations: Features with negative correlations (indicated by cooler colors, such as blue) move in opposite directions. When one attribute increases, the other tends to decrease.
- Strong Correlations: Look for bright red or dark blue squares. These represent strong correlations. For instance, if pH and hardness have a dark blue square, it means they are negatively correlated.
- Weak Correlations: Lighter colors (closer to white) indicate weaker correlations. These might not be as significant for your analysis.

The heatmap can help to identify which features are most relevant for determining the type of migraine. For example, features associated with "Migraine Without Aura" type:

- Moderate to severe head pain: Usually on one side of the head.
- Pulsing or throbbing pain: Common migraine headaches often have this characteristic.
- Sensitivity to light and/or noise: Photophobia and phonophobia are common symptoms.
- Nausea and/or vomiting: Gastrointestinal symptoms are often present.

2.3 Data pre-processing

Once data is collected, it must undergo pre-processing to be suitable for machine learning analysis. This pre-processing may involve various tasks, such as removing missing or outlier values, normalizing variables, and transforming categorical data into numerical variables. These steps are crucial to ensure accurate and reliable results from machine learning models.

In our specific scenario, the dataset doesn't contain any missing values or outliers. Our primary tasks involve:

- **Converting the "Type" Feature:**

We need to convert the categorical feature "Type" into a numeric representation. This transformation allows us to work with machine learning algorithms that require numerical input. Depending on the specific types of migraine (e.g., "Migraine with Aura," "Migraine without Aura," etc.), we'll assign numeric labels (e.g., 0, 1, 2, etc.) to each type.

- **Removing the "Ataxia" Column:**

Since the “Ataxia” column has the same value (0) for all rows, it doesn’t provide any useful information for our analysis. We can safely delete this column from our dataset to simplify our feature set.

2.4 Predictive analysis

Once data has been analyzed, cleaned, and features have been transformed into usable formats, the subsequent stage involves selecting, constructing, training, and testing machine learning models.

2.4.1Machine learning method

This section presents our first proposition.

- **Models’ selection**

For our model selection process, we propose utilizing various algorithms commonly employed in similar studies found in the literature. We have chosen several benchmark models for comparison, which include: Random Forest, logistic regression, KNN, SVM, Naïve Bayes, Decision tree, XGboost

Furthermore, in addition to the aforementioned algorithms, we suggest incorporating a recent addition from the gradient boosting family: the XGBoost algorithm. Its introduction will be detailed in the subsequent section.

Extreme Gradient boosting:

Extreme Gradient boost also known as XGBoost is a highly efficient and scalable implementation of the gradient boosting package. The XGBoost offers better scalability and can perform parallel and distributed computing offering faster learning and model exploration. This newly introduced package offers various objective functions including Ranking, Classifier, and Regression. With its advantage of adjustable parameters, XGBoost can make predictions on the set of defined standard features and offers better results than other machine learning models. [39]

- **Data Division**

In machine learning, it is common practice to divide available data into two sets - the training set and the testing set. The purpose of this division is to ensure that the model being developed is not overfitting or underfitting the data. The training set is utilized to train the prediction model, while the test set is used to evaluate the model's performance on new or unseen data. This approach helps in developing a more accurate and reliable model that can generalize well to new data.

- **Results**

Model	Accuracy
Random forest	85,83
Logistic regression	89,16
KNeighbors	74,16

SVM	89,16
Naïve Bayes	86,66
Decision tree	84,16
XGboost	90

```
Logistic Regression: 89.16666666666667
K Nearest neighbors: 74.16666666666667
Support Vector Classifier: 89.16666666666667
Naive Bayes: 86.66666666666667
Decision tree: 84.16666666666667
Random Forest: 85.83333333333333
```

Figure 23: models' accuracy.

```
Accuracy: 0.90
              precision    recall  f1-score   support

           0       0.75      0.50      0.60         6
           1       0.50      0.67      0.57         3
           2       0.93      1.00      0.96        13
           3       1.00      0.75      0.86         4
           4       0.00      0.00      0.00         2
           5       0.94      0.98      0.96        49
           6       1.00      1.00      1.00         3

    accuracy                           0.90        80
   macro avg       0.73      0.70      0.71        80
weighted avg       0.89      0.90      0.89        80
```

Figure 24: XGboost accuracy.

- **Discussion and comparison:**

In our final debate, it's important to address related works, particularly those that utilize the same dataset. We can compare the results obtained in this study with the results from our study using different models.

In our study, we used seven different models: **LR**, **DT**, **RF**, **SVM**, **KNN**, **XGboost**, **Naive Bayes**

Among these models, **XGboost performed the best**, with a performance of **0.90**. This indicates that our XGboost model accurately predicted the target values. Comparatively, the work [125] used five models, **NN**, **KNN**, **SVM**, **LR**, and **DT**, to analyze the same dataset. They obtained a Logistic Regression Model Score of 0.95. Although these results are also promising, they are slightly larger than those in our study.

In conclusion, our study demonstrated that the **XGBoost** model outperformed the six other models we utilized. These results are consistent with those of other studies using similar models. However, further research is necessary to improve the performance of the models and to gain a deeper understanding of the factors influencing their results. Below is a comparison table between our work and other studies, based on the dataset used, the models employed, the evaluation metrics, and the results obtained:

	Dataset used	**Models used**	**Results**
Our work	Migraine classification dataset	LR, DT, RF, SVM, KNN, XGboost, Naive Bayes	Best result XGboost 0,90
[53]	Migraine classification dataset	NN, KNN, SVM, LR,and DT	Best result Logistic regression **0.95**

Table 06: Comparing the results of jobs using the same Dataset with our work

2.4.2. Deep learning method

This section presents our second proposition.

- **Tabular-to-Image Transformation and CNN architecture:**

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly well-suited for image processing tasks. Unlike traditional neural networks, CNNs leverage the spatial structure of data through the use of convolutional layers, which apply filters to input data to capture local patterns such as edges, textures, and other features. These patterns are then combined and abstracted through pooling layers, which reduce the dimensionality of the data while preserving important features, ultimately enabling the network to learn complex representations.

The second proposition aims to innovate in the field of migraine classification by leveraging the powerful capabilities of Convolutional Neural Networks (CNNs). Traditional machine learning approaches, while effective, may not fully exploit the complex patterns present in tabular data. This proposition seeks to transform tabular data into image-like representations, thereby enabling the application of CNNs to this domain.

In this research, the CNN architecture is adapted to handle tabular data transformed into image-like structures, thereby harnessing the powerful feature extraction capabilities of CNNs to enhance the classification accuracy of migraine types. The diagram below illustrates the detailed architecture of the CNN model used in this study, highlighting the specific layers and their configurations.

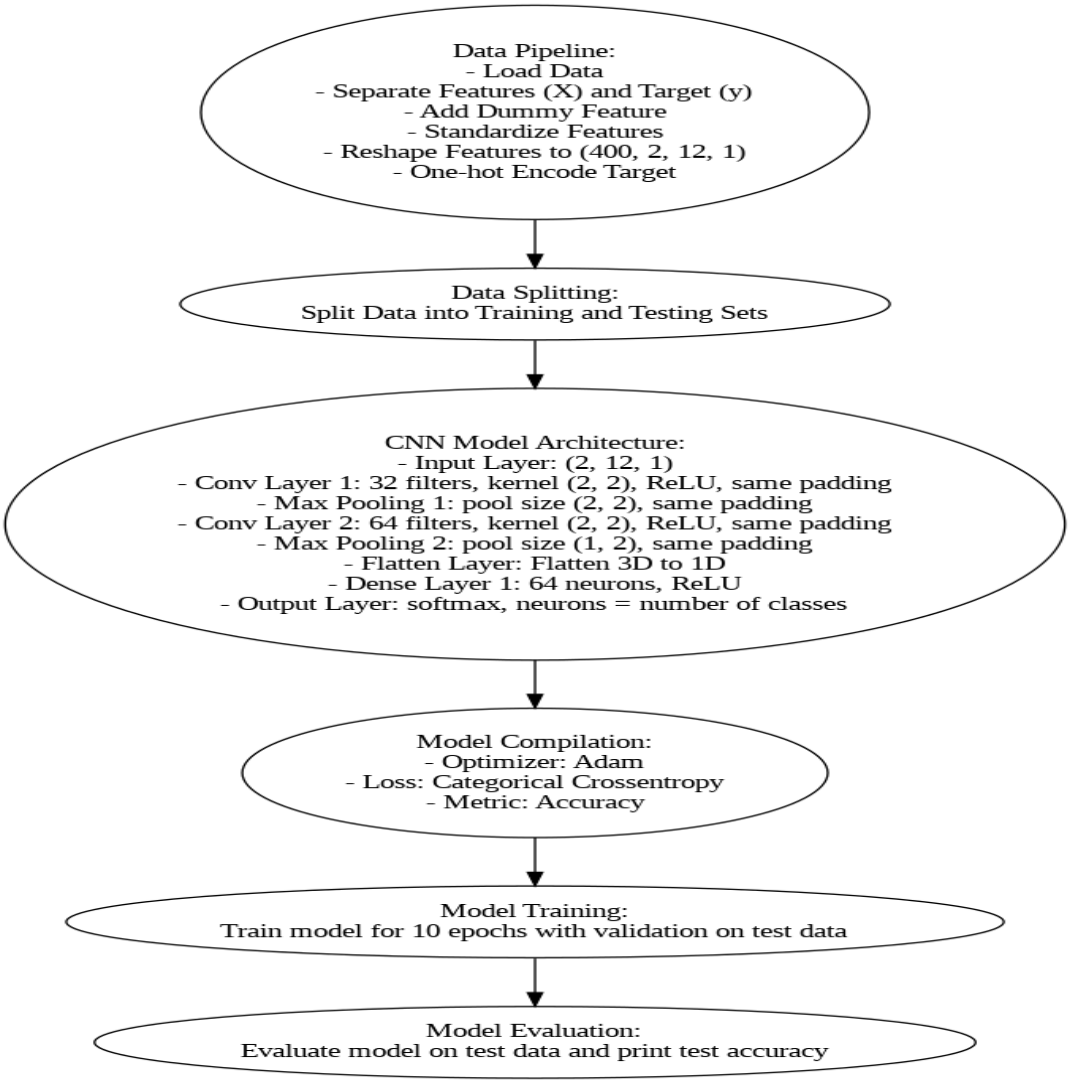


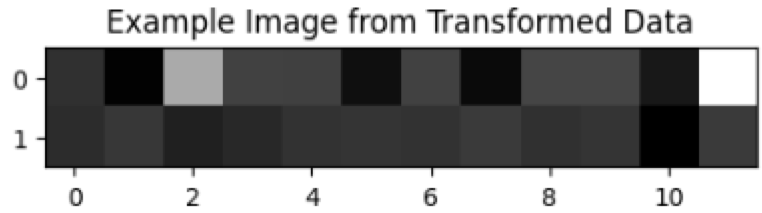
Figure 25: CNN diagram**.**

- **Description:**

We have followed a structured approach for building and evaluating a CNN model for predicting types of migraine. Here's a summary of the key steps:

Data Pipeline:

- We separated the features (X) and the target (y).
- We added a dummy feature and standardized the features.
- We reshaped the features into 2D images (400, 2, 12, 1).



- We one-hot encoded the target variable.

Data Splitting:

- We split the data into training and testing sets.

CNN Model Architecture:

- We defined the input layer as (2, 12, 1).
- We created convolutional layer 1 with 32 filters, a kernel size of (2, 2), ReLU activation, and the same padding.
- We added max pooling layer 1 with a pool size of (2, 2) and same padding.
- We built convolutional layer 2 with 64 filters, a kernel size of (2, 2), ReLU activation, and the same padding.
- We included max pooling layer 2 with a pool size of (1, 2) and same padding.
- We flattened the 2D tensor into 1D.
- We implemented dense layer 1 with 64 neurons and ReLU activation.
- We created the output layer with Softmax activation and neurons equal to the number of classes.

Model Evaluation:

- We evaluated the model on the test data.
- We calculated the test accuracy.

Our approach seemed reasonable for building a CNN model to classify different types of migraine. The data preprocessing steps, such as adding a dummy feature, standardizing, and reshaping, were essential for preparing the data for the CNN model. The model architecture included convolutional layers for learning spatial features, max-pooling layers for dimensionality reduction, and dense layers for classification.

- **Comparison with the first proposition:**

In our first proposition, we explored the use of seven different models for migraine-type prediction: Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), XGBoost, and Naive Bayes. Among these, the XGBoost model performed the best with an accuracy of 90%, indicating its robustness in accurately predicting the target values. Comparatively, related work using the same dataset and employing five models (Neural Networks (NN), KNN, SVM, LR, and DT) achieved a higher Logistic Regression score of 95%. Despite the promising results of our XGBoost model, it fell slightly short of the Logistic Regression model used in the comparative study.

In contrast, our second proposition utilized a Convolutional Neural Network (CNN) for predicting migraine types and achieved an impressive accuracy of 93%. The CNN's architecture, consisting of convolutional and pooling layers followed by fully connected layers, allows for effective extraction and integration of relevant features, leading to high predictive accuracy.

Aspect	First proposition	Second proposition
Models used	LR, DT, RF, SVM, KNN, XGBoost, Naive Bayes	CNN
Best performing model	XGBoost	CNN
Accuracy	90%	93%
Data utilization	Migraine classification dataset	Migraine classification dataset
Model complexity	Lower (traditional machine learning models)	Higher (deep learning model)
Computational Requirements	Moderate	High
Need for further research	Improve model performance, understand influencing factors	Same, but also optimizes computational efficiency

Table 07: comparison between the first proposition and the second.

Conclusion

In conclusion to this chapter, we utilized the powerful tools of the Python language to conduct our study. We worked with a specific dataset and followed several key steps to predict migraine types using artificial intelligence. Our architectural contribution, illustrated by a diagram, demonstrated our methodical approach and in-depth understanding of the problem. The code screenshots provided a tangible representation of our work. Through this combination of tools and methods, we were able to make accurate predictions of migraine types, laying the groundwork for future improvements in healthcare. Additionally, we discussed two proposition models: the XGBoost model and the CNN model. The XGBoost model achieved a commendable accuracy of 90%, while the CNN model demonstrated an even higher accuracy of 93%. These comparisons highlighted the potential and effectiveness of advanced machine learning and deep learning techniques in improving migraine prediction accuracy.

General Conclusion

Summary of Research

The following is a summary of the content:

In this thesis, our research focused on implementing advanced machine learning and deep learning techniques to enhance the accuracy of migraine-type classification using a comprehensive dataset. We primarily explored two approaches:

1. Ensemble Learning with XGBoost: We utilized the XGBoost algorithm to classify migraine types and found that it outperformed traditional machine learning models in terms of accuracy and efficiency.

2. Tabular-to-Image Transformation and CNN: We developed a new method that involves transforming tabular migraine data into images, thus enabling the use of CNNs for classification. This approach yielded remarkable accuracy, surpassing traditional and ensemble learning models

Contributions:

- **Descriptive, Diagnostic, and Predictive Analysis:** We provided a comprehensive framework for analyzing migraine data, integrating descriptive, diagnostic, and predictive analyses to offer a holistic view of the dataset.
- **Novel Tabular-to-Image Transformation**: We introduced a new method for transforming tabular data into images, enabling the use of CNNs for classification tasks.
- **Enhanced Classification Accuracy:** The effectiveness of advanced machine learning and deep learning techniques in improving the accuracy of migraine-type classification was demonstrated.

Implications:

Our findings have significant implications for medical informatics and migraine management. By enhancing the accuracy and reliability of migraine classification, healthcare providers can develop more targeted and effective treatment plans, ultimately improving patient outcomes. Additionally, the novel methodologies introduced in this thesis can be adapted and applied to other medical classification problems, further extending their impact.

Future Work:

Future research could explore the integration of real-time data and the application of more advanced deep learning architectures to further enhance classification accuracy. Additionally, investigating the use of these techniques in other medical conditions could provide valuable insights and advancements in the field.

References

[1] Runkler, T. A. (2020). Data analytics. Springer Fachmedien Wiesbaden.

[2] Soltanpoor, R., & Sellis, T. (2016, September). Prescriptive analytics for big data. In Australasian database conference (pp. 245-256). Springer, Cham.

[3] Rajaraman, V. (2016). Big data analytics. Resonance, 21(8), 695-716.

[4] Network-Based Visual Analysis of Tabular Data Zhicheng Liu∗ Shamkant B. Navathe† John T. Stasko

[5] Deep Neural Networks and Tabular Data: A Survey Vadim Borisov , Tobias Leemann , Kathrin Seßler , Johannes Haug , Martin Pawelczyk , and Gjergji Kasnec

[6] Delen, D., & Demirkan, H. (2013). Data, information, and analytics as services. Decision Support Systems, 55(1), 359-363.

[7] Rajaraman, V. (2016). Big data analytics. Resonance, 21(8), 695-716.

[8] Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A framework for pandemic prediction using big data analytics. Big Data Research, 25, 100190.

[9] Soltanpoor, R., & Sellis, T. (2016, September). Prescriptive analytics for big data. In Australasian database conference (pp. 245-256). Springer, Cham.

[10] Eckerson, W. W. (2007). Predictive analytics. Extending the Value of Your Data

Warehousing Investment. TDWI Best Practices Report, 1, 1-36.

[11] Soltanpoor, R., & Sellis, T. (2016, September). Prescriptive analytics for big data. In Australasian database conference (pp. 245-256). Springer, Cham.

[12] Poornima, S., & Pushpalatha, M. (2020). A survey on various applications of prescriptive analytics. International Journal of Intelligent Networks, 1, 76-84.

[13] Yao, X., & Liu, Y. (2014). Machine learning. In Search Methodologies (pp. 477-517). Springer, Boston, MA.

[14] El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In machine learning in radiation oncology (pp. 3-11). Springer, Cham.

[15] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[16] Huang, J. C., Ko, K. M., Shu, M. H., &amp ; Hsu, B. M. (2020). Application and comparison of several machine learning algorithms and their integration models in regression problems. Neural Computing and Applications, 32(10), 5461-5469.

[17] CLASSIFYING THE SUPERVISED MACHINE LEARNING AND COMPARING THE PERFORMANCES OF THE ALGORITHMS Rathindra Nath Mohalder1, Dr. Md. Alam Hossain2, and Nazmul Hossain

[18] Yifei Huang • Yuhua Liu • Chenhui Li • Changbo Wang GBRTVis: online analysis of gradient boosting regression tree

[19] Mahesh, B. (2020). Machine Learning Algorithms-A Review. International Journal of Science and Research (IJSR). [Internet], 9, 381-386.

[20] Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., ... & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications, and research challenges. IEEE Access, 7, 65579-65615.

[21] Comparison of K-Means, K-Medoids, and K-Means++algorithms based on the Calinski-Harabasz index for COVID-19 epidemic in China Zhengcao Hu Taiyuan University of Technology, No. 209, University City, Yuci District, Jinzhong City, Shanxi Province, China.

[22] ADAPTIVE DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE (DBSCAN) ACCORDING TO DATA WEI-TUNG WANG', YI-LEH WU', CHENG-YUAN TANG, MAW-KAE HOR[3]

[23] https://www.geeksforgeeks.org/hierarchical-clustering-in-data-mining/

[24] Data Analytics in Healthcare: A Tertiary Study Toni Taipalus1 · Ville Isomöttönen1 · Hanna Erkkilä1 · Sami Äyrämö1

[25] Intelligent health data analytics: A convergence of artificial intelligence and big data Syed Sibte Raza Abidi, PhD1 and Samina Raza Abidi, PhD, MBBS2

[26] Anand J. Kulkarni · Patrick Siarry · Pramod Kumar Singh · Ajith Abraham · Mengjie Zhang · Albert Zomaya · Fazle Baki Editors Big Data Analytics in Healthcare

[27] Functional connectivity of the language area in migraine: a preliminary classification model Chen Gou1,2†, Shuangfeng Yang1,2†, Qianmei Hou1,2, Peter Rudder3, Onur Tanglay3, Isabella Young3, Tingting Peng1,2, Weiwei He1,2, Liuyi Yang4, Karol Osipowicz3, Stephane Doyen3, Negar Mansouri3, Michael E. Sughrue3* and Xiaoming Wang1,2

[28] https://www.ncbi.nlm.nih.gov/books/NBK513302/

[29] https://www.mayoclinic.org/diseases-conditions/migraine-with-aura/symptoms-causes/syc-20352072

[30] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4135274/

[31] https://www.medicalnewstoday.com/articles/320013

[32] https://my.clevelandclinic.org/health/diseases/silent-migraine

[33] https://pubmed.ncbi.nlm.nih.gov/15566415/

[34] What We Gain from Machine Learning Studies in Headache Patients Roberta Messina and Massimo Filippi 1 Neuroimaging Research Unit, Division of Neuroscience, Institute of Experimental Neurology, Milan, Italy, 2 Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy, 3 Vita-Salute San Raffaele University, Milan, Italy

[35] Development and Internal Validation of a Multivariable Prediction Model for Individual Episodic Migraine Attacks Based on Daily Trigger Exposures Katherine K. Holsteen, PhD; Michael Hittle, BS; Meredith Barad, MD; Lorene M. Nelson, PhD, MS

[36] The Clinical Relevance of Artificial Intelligence in Migraine Angelo Torrente 1, Simona Maccora 1,2, Francesco Prinzi 1,3, Paolo Alonge 1, Laura Pilati 1,4, Antonino Lupica 1, Vincenzo Di Stefano 1, Cecilia Camarda 1, Salvatore Vitabile 1 and Filippo Brighina 1

[37] ENHANCING MIGRAINE DIAGNOSIS: A SYMPTOM-BASED CLASSIFICATION APPROACH WITH SVM AND DECISION TREE LAMA SADI AWAD 1, ENAS TAWFIQ NAFFAR 2 and SALEH AL-SHARAEH 3 1 King Abdullah II School of Information Technology, The University of Jordan Amma. Email: lma9220481@ju.edu.jo 2 Department of Software Engineering, Philadelphia University. Email: enaffar@philadelphia.edu.jo 3 Professor, King Abdullah II School of Information Technology, The University of Jordan Amma.

[38] Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data Yolanda Garcia-Chimeno1,2*, Begonya Garcia-Zapirain1,2, Marian Gomez-Beldarrain3, Begonya Fernandez-Ruanova3, and Juan Carlos Garcia-Monco

[39] Comparative investigation of machine learning algorithms for detection of epileptic seizures Akash Sharma∗, Neeraj Kumar, Ayush Kumar, Karan Dikshit, Kusum Tharani and Bharat Singh Electrical and Electronics Engineering Department, Bharati Vidyapeeth's College of Engineering, New Delhi, India

[40] Migraine with aura detection and subtype classification using machine learning algorithms and morphometric magnetic resonance imaging data Katarina Mitrović1*, Igor Petrušić2, Aleksandra Radojičić3,4, Marko Daković2, and Andrej Savić5

[41] Automatic migraine classification using artificial neural networks [version 2; peer review: 1 approved, 2 approved with reservations] Paola A. Sanchez-Sanchez, José Rafael García-González, Juan Manuel Rúa Ascar

[42] Validation of an algorithm for automated classification of migraine and tension-type headache attacks in an electronic headache diary Aaron Roesch1*, Markus A Dahlem2, Lars Neeb3, and Tobias Kurth

[43] Computerized migraine diagnostic tools: a systematic review Yohannes W. and Robert P. Cowan2

[44] MyGraine: Predicting Migraines Through Various Machine Learning Models Utilizing User-Inputted Data Rebecca S. Zhu, Rucha Dave Nashua High School South, 36 Riverside Street, Nashua, NH, 03062, U.S.A.

[45] Migraine, Tension-Type, and Cluster-Type of Headaches Classification by Using Immunos Algorithms Article in Journal of Medical Imaging and Health Informatics ·

[46] Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data Yolanda Garcia-Chimeno1,2*, Begonya Garcia-Zapirain1,2, Marian Gomez-Beldarrain3, Begonya Fernandez-Ruanova3, and Juan Carlos Garcia-Monco4

[47] To what extent are patients with migraine able to predict attacks? Ana B Gago-Veiga1 Josué Pagán2,3 Kevin Henares2,4 Patricia Heredia1 Nuria González-García5 María-Irene De Orbe4 Jose L Ayala3,4 Mónica Sobrado1 Jose Vivancos1

[48] Machine learning approach to predict medication overuse in migraine patients Patrizia Ferroni a, b,1, Fabio M. Zanzotto c,1, Noemi Scarpato b, Antonella Spila a, Luisa Fofi d, Gabriella Egeo d, Alessandro Rullo e, Raffaele Palmirotta f, Piero Barbanti b, d,2, Fiorella Guadagni

[49] Classification of multi-channel EEG signals for migraine detection Selahaddin Batuhan Akben Osmaniye Korkut Ata University Deniz Tuncel Kahramanmaras Sutcu Imam University Ahmet Alkan Kahramanmaras Sutcu Imam University

[50] Improving Migraine Diagnosis by Using Binary and Multiclass Classification of Machine Learning (liacs.nl)

[51] A High Performance System for the Diagnosis of Headache via Hybrid Machine Learning Model (thesai.org)

[52] Applying Machine Learning Models to Diagnosing Migraines with EEG Diverse Algorithms (anapub.co.ke)

[53] https://codeocean.com/capsule/1269964/tree/v1