electronics

Review

Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection

Mohammad Moshawrab [1,*], Mehdi Adda [1], Abdenour Bouzouane [2], Hussein Ibrahim [3,*] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
[*] Correspondence: mohammad.moshawrab@uqar.ca (M.M.); hussein.ibrahim@itmi.ca (H.I.); Tel.: +1-(581)624-9394 (M.M.)

Abstract: Machine Learning (ML) and Deep Learning (DL) are derivatives of Artificial Intelligence (AI) that have already demonstrated their effectiveness in a variety of domains, including healthcare, where they are now routinely integrated into patients' daily activities. On the other hand, data heterogeneity has long been a key obstacle in AI, ML and DL. Here, Multimodal Machine Learning (Multimodal ML) has emerged as a method that enables the training of complex ML and DL models that use heterogeneous data in their learning process. In addition, Multimodal ML enables the integration of multiple models in the search for a single, comprehensive solution to a complex problem. In this review, the technical aspects of Multimodal ML are discussed, including a definition of the technology and its technical underpinnings, especially data fusion. It also outlines the differences between this technology and others, such as Ensemble Learning, as well as the various workflows that can be followed in Multimodal ML. In addition, this article examines in depth the use of Multimodal ML in the detection and prediction of Cardiovascular Diseases, highlighting the results obtained so far and the possible starting points for improving its use in the aforementioned field. Finally, a number of the most common problems hindering the development of this technology and potential solutions that could be pursued in future studies are outlined.

Keywords: multimodal machine learning; multimodal learning; data heterogeneity; data fusion; model heterogeneity; model fusion; diseases prediction; cardiovascular diseases; Internet of Things; smart wearables

check for updates

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection. *Electronics* **2023**, *12*, 1558. https://doi.org/10.3390/electronics12071558

Academic Editors: Francisco Luna-Perejón, Lourdes Miró Amarante and Francisco Gómez-Rodríguez

Received: 2 March 2023
Revised: 14 March 2023
Accepted: 25 March 2023
Published: 26 March 2023

1. Introduction

Artificial Intelligence (AI) has experienced rapid growth over the past two decades. The concept of AI has been around since 1950, and the term itself was coined in 1965 at the Dartmouth Summer Workshop, which is considered the founding event of AI as a field [1]. However, the growth in Information and Communication Technologies (ICTs) and the increasing power of computers have contributed significantly to the increasing feasibility and adoption of AI [2]. AI technologies are becoming more advanced and are capable of analyzing enormous amounts of data, learning from past experiences, and making predictions based on patterns and trends [3]. Despite the popularity of AI, there is no single definition for this technology. Researchers in [4], for example, defined it as a set of tools and techniques that use principles and devices from various fields, such as computation, mathematics, logic, and biology, to address the problem of realizing, modeling, and mimicking human intelligence and cognitive processes. Furthermore, the authors define in [5] AI as the study of an "Intelligent Agent", i.e., machines that are able to recognize and understand their environment and consequently take appropriate actions to increase their chances of achieving their goals. In an attempt to unify definitions, the authors defined in [6] AI as a program that can cope in an arbitrary world no worse than a human. These different definitions reflect the different competencies of AI, which explains the diversity of AI implementations in our daily lives.

Machine Learning (ML) [7], Deep Learning (DL) [8], Federated Machine Learning (FL) [9], and Multimodal Machine Learning [10] are all well-known and popular derivatives of the AI concept that have been adopted by users and applied in various aspects of our daily lives. These different branches of AI are depicted in Figure 1. In this context, Machine Learning is defined as a field of study that focuses on the development of algorithms and statistical models that enable computer systems to learn from data and make predictions or decisions without being explicitly programmed. It involves the application of various approaches, such as supervised and unsupervised learning, Reinforcement Learning, and Deep Learning, that allow computers to automatically improve their performance on a given task through experience [7].

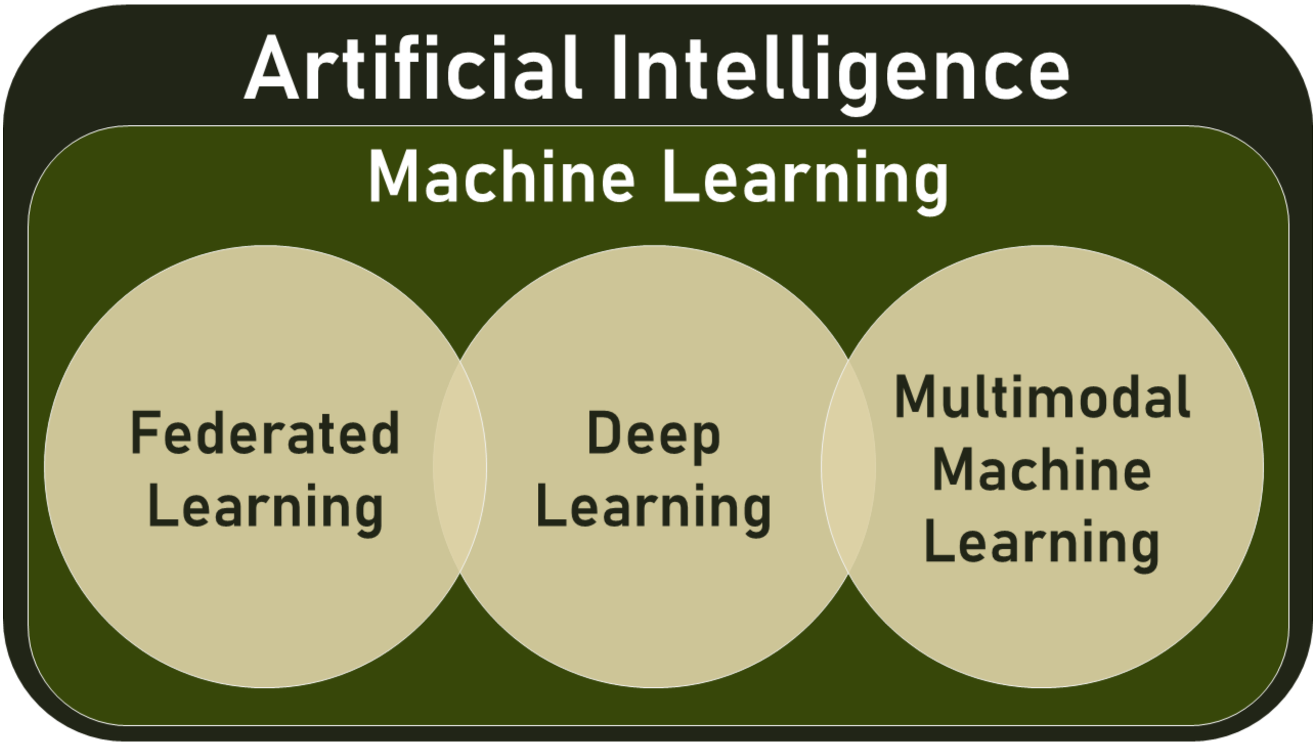


Figure 1. Artificial intelligence branches.

On the other hand, Machine Learning has demonstrated high efficiency in solving classification and regression problems. Machine Learning's ability to extract meaningful insights and patterns from vast and complicated datasets and use this knowledge to make accurate predictions, automate decision making, and enable intelligent systems to learn and adapt in real-time is fundamental to its success. This success has led researchers from different fields to implement ML algorithms, and their efficiency can be observed in various fields, such as:

- Healthcare services [11–13];
- Image, speech and pattern recognition [14,15];
- Internet of Things (IoT) and smart cities [14,16];
- Cybersecurity and threat intelligence [17];
- Natural language processing and sentiment analysis [18];
- User behavior analytics and context-aware smartphone applications [14,15];
- E-commerce and product recommendations [14,15];
- Sustainable agriculture [19];

- Industrial applications [20].

1.1. Machine Learning Domain Challenges

The great success of Machine Learning is not magic but the result of its ability to analyze large amounts of data at high speed and with high accuracy. However, the field of ML still suffers from various challenges and obstacles arising from different problems. Table 1 below summarizes the Machine Learning challenges and categorizes them based on their source. These challenges have been extensively studied in the literature, and more details can be found in several articles, such as [9,21–23].

Table 1. Machine Learning domain common challenges.

Group	Challenges	
Data-Related Challenges [21,22]	Data Availability and Accessibility [23] Data Locality [16]	
	Data Readiness [23]	Data Heterogeneity Noise and Signal Artifacts Missing Data Classes Imbalance
	Data Volume	Course of Dimensionality Bonferroni principle [24]
	Feature Representation and Selection	
Models Related Challenges [25,26]	Accuracy and Performance Model Evaluation Variance and Bias Explainability	
Implementation-Related Challenges [23,27]	Real-Time Processing Model Selection Execution Time and Complexity	
General Challenges [25,26]	User Data Privacy and Confidentiality User Technology Adoption and Engagement Ethical Constraints	

1.2. Heterogeneity: Motivation(s) behind Multimodal ML

Advances in sensor technologies, storage concepts, communication networks, and other tools have driven data collection [28]. According to recent figures from Statista [29], the total amount of data generated worldwide will reach 64.2 zettabytes or 6.42×10^{16} Megabytes in 2020. This increase exceeded predictions due to increasing demand as a result of the COVID-19 pandemic, as more individuals worked and studied from home and increasingly used utilized home entertainment alternatives. For the above reasons, data volumes are expected to reach 180 zettabytes in the next five years by 2025.

However, these data differ in type, structure, format, usability, lifespan, and other aspects. This heterogeneity poses several challenges in Machine Learning, as it can make it difficult to use data from different resources to gain useful insights or build accurate models. There are many types of heterogeneity, the most common of which are listed below [21,30,31]:

- Structured vs. unstructured data: structured data are highly organized and usually follow a specific schema, while unstructured data have no predefined structure or organization;
- Numeric vs. categorical data: Numeric data are quantitative and can be expressed as numbers, while categorical data are qualitative and represent discrete values, such as colors, types, or labels;
- Temporal data: This type of data contains time-stamped information and can be used to analyze patterns and trends over time;

- Multimodal data: This type of data combines different types of information, such as text, audio, images, and videos.

Thus, dealing with heterogeneous data requires careful processing and feature engineering to put the data into the form required for a single Machine Learning model [31]. In addition, multiple preprocessing steps may be required to analyze heterogeneous data, such as normalization, scaling, or other steps. In some cases, however, it may seem impossible to analyze heterogeneous data, even though training the model with this variety of resources improves its feasibility and increases confidence in its predictions.

For example, Magnetic Resonance Imaging (MRI) analysis using ML models has shown high efficiency in predicting Cardiovascular Diseases (CVDs), as shown in [32]. In addition, smart wearables equipped with ML models are also highly feasible in predicting cardiac disease, as shown in [33]. In addition, the use of Electronic Health Records (EHRs) collected from various health centers such as clinics, hospitals, or smart homes is also a good source for Cardiovascular Disease prediction using ML algorithms [34]. However, trying to merge these three types of data seems to be technically impossible because the first data source, namely MRI images, are stored in the form of medical electronic image files, and the data collected by wearables are structured data, while EHRs can be a collection of both structured and unstructured data, free text reports, medical examination data, or other formats. In the real world, a physician may analyze all of these data to make a more accurate diagnosis, though it is not easy to analyze these data sets simultaneously using the same model. This case is illustrated in Figure 2 below.

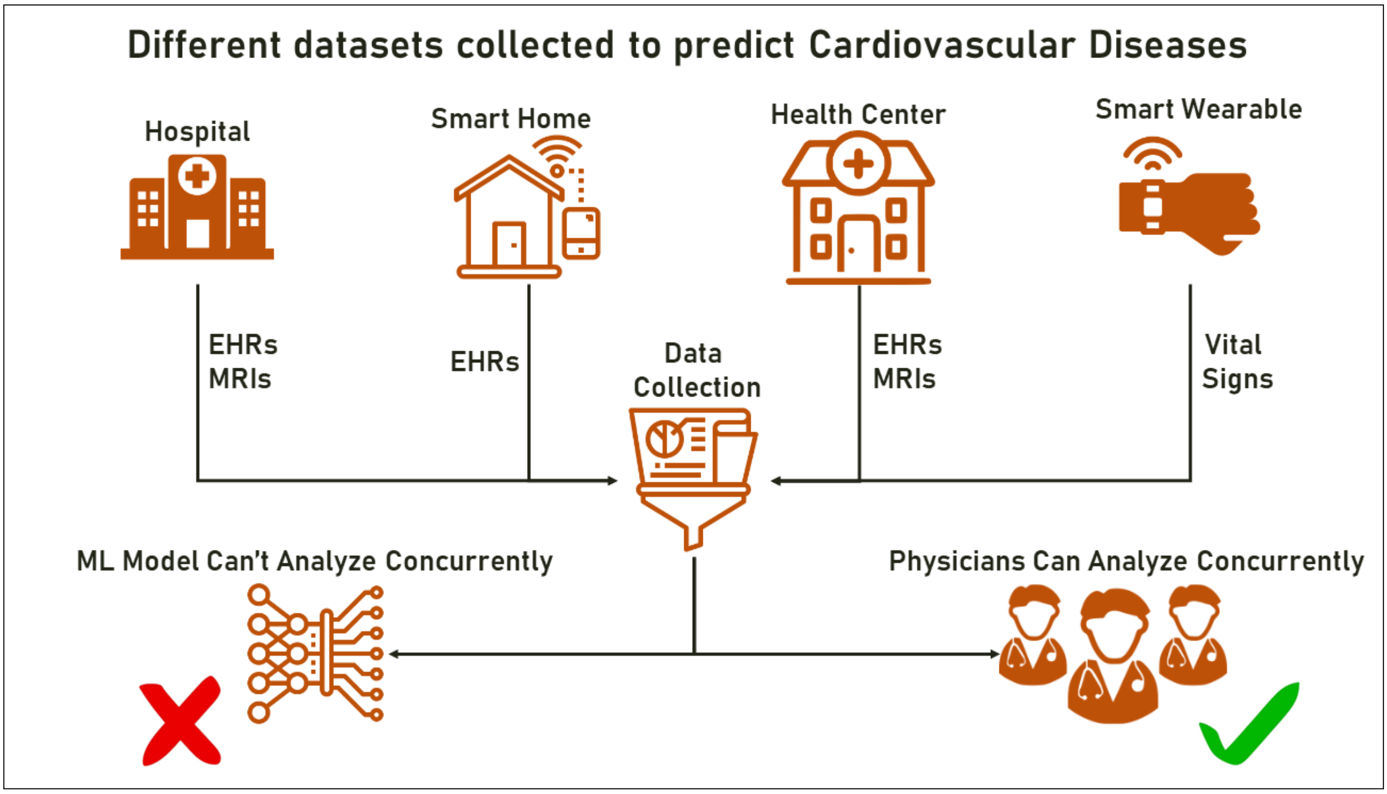

Different datasets collected to predict Cardiovascular Diseases
Hospital
Smart Home
Health Center
Smart Wearable
EHRs MRIs
EHRs
Data Collection
EHRs MRIs
Vital Signs
ML Model Can't Analyze Concurrently
Physicians Can Analyze Concurrently

Figure 2. Prediction of CVDs with heterogeneous data—a showcase.

In this context, Multimodal Machine Learning is proposed as a solution to the challenge of data heterogeneity in ML. Multimodal ML gives models the ability to analyze different data within the same ML workflow, whether by merging different datasets, by merging different models, or both, to arrive at a single result, such as the diagnosis of CVDs in the showcase mentioned above [10]. The ability to analyze these heterogeneous data with multiple views can be of varying importance to a learning task. Therefore, merging all of these data sets and treating them with equal importance is unlikely to lead to optimal learning outcomes [30].

1.3. Machine Learning and Healthcare

The importance of health to human life cannot be overstated, as it is essential for meeting basic needs, pursuing goals, maintaining relationships, and having an adequate quality of life, and poor health can have significant financial and societal consequences. Therefore, researchers are constantly striving to improve the quality of healthcare services. In this context, Artificial Intelligence and its branches, such as Machine Learning and Deep Learning, have been incorporated into healthcare services due to their high feasibility and usability in this field. Machine learning, in particular, is a powerful tool that has the potential to revolutionize healthcare in many ways [35]. ML has made remarkable progress in healthcare, not because of any mystical powers, but because of its superior data processing capabilities compared to those of humans. Because of its speed and precision, thousands of AI applications have already been developed for healthcare, making it a potentially revolutionary tool for solving a wide range of healthcare problems [36].

Machine Learning has been used in various areas of healthcare. Whether diagnosing diseases or even predicting diseases, it has proven to be very useful. Moreover, the development of communication tools, such as smart wearables equipped with Machine Learning and Deep Learning models, has opened the door to real-time continuous monitoring. In this context, smart wearables have shown high feasibility in predicting various diseases such as Cardiovascular Diseases [33], diabetes [37], liver disease [38], fatigue and stress [39], mental illness [40], and many other diseases [41]. In addition, ML models have been used to increase the efficiency of healthcare decision systems [42]. In addition, ML has also been used in the field of genomic medicine [43]. Overall, ML has succeeded in transforming health services and creating personalized digital health services that support physicians and improve the overall quality of public health [44].

Therefore, considering the importance of healthcare, it is urgent to improve the efficiency of ML. The use of state-of-the-art methods and the removal of obstacles to progress are essential to improving performance. The challenges described previously are reflected in the barriers to expanding the use of ML in healthcare, which are common to all ML implementations across all diseases. With this in mind, new solutions that could help promote the use of ML will lead to improved applications in a variety of settings.

1. Define the scope of the review: Clearly define the scope and objective of the review article. What is the main topic or research question that the review aims to address? What specific subtopics or themes will be covered? 2. Identify the key concepts and themes: Based on the scope and objective of the review, identify the key concepts and themes that will be discussed in the article. These should be organized in a logical and coherent manner that supports the overall objective of the review. 3. Develop a framework for presenting the review: Once the key concepts and themes have been identified, develop a framework for presenting the review. This could involve organizing the content chronologically, thematically, or conceptually, depending on the nature of the review and the key concepts and themes identified. 4. Clearly articulate the review framework: Finally, clearly articulate the review framework in the introduction or early sections of the review article. This will help to orient readers to the overall structure and organization of the review and make it easier for them to follow the content. Overall, the goal is to provide a clear and structured overview of the review article that highlights the key concepts and themes and guides the reader through the content in a logical and coherent manner.

1.4. Review Framework: Scope, Outline and Main Contributions

In this article, Multimodal Machine Learning is explored, and its role as a solution to the challenge of heterogeneity is detailed. In addition, the use of Multimodal ML in

Cardiovascular Disease detection and prediction is technically reviewed to support its use in this field.

1.4.1. Scope of Research

To achieve the objectives of the study, Multimodal Machine Learning has been explored, along with the data fusion concept, which is the basis of the technology under study. In addition, the technical perspectives of Multimodal ML are studied, and the workflows related to it are examined. Furthermore, a comparison between Multimodal ML and other known techniques is made in order to distinguish between these different techniques. On the other hand, distinct areas where Multimodal ML is used are inspected, and a comprehensive overview of its application in Cardiovascular Diseases, including the state of the art, is therefore obtained. In addition, these implementations were analyzed from different perspectives to understand the limitations and future areas of research. Finally, the challenges and future recommendations associated with advancing this field are reviewed.

1.4.2. Research Questions

The scope of the article defined in the previous section is summarized by the research questions mentioned in the list below:

- What is Multimodal Machine Learning?
- What are the motivations for this technology?
- What are the technical perspectives on which Multimodal ML is based?
- What are the differences between Multimodal ML, classical ML, Multimodal datasets, ensemble ML and other techniques?
- What are the existing Multimodal ML frameworks, and what contributions do each make?
- What is the state of the art in the use of Multimodal ML in CVD prediction, and what technical summaries can be derived?
- What challenges still impede progress in this area, and what approaches could be taken to overcome these issues?

1.4.3. Outline

To answer the above questions, the article is outlined as follows. In Section 2, Multimodal ML is reviewed from various angles, including technical definition(s), differences from other domains, such as classical ML, ensemble ML and others, available frameworks, and other details. Then, in Section 3, the use of Multimodal ML technology in CVD detection and prediction is presented by listing the state of the art in this field and discussing the technical details of the implementations mentioned in the literature. Later, in Section 4, the challenges that hinder progress in this field are discussed, and therefore, some future perspectives that could help in overcoming these challenges are proposed. This article attempts to answer the following questions:

1.4.4. Comparison with Previous Review Frameworks

The topic of Multimodal ML has been a hot and trending topic in recent years. As a result, numerous studies have already addressed this topic, with a large proportion of these studies reviewing Multimodal ML. However, this article proposes several new ideas that add to the knowledge of Multimodal ML. First, this study proposes a technical study for Multimodal ML that, on the one hand, helps to understand this technology and distinguish it from other existing AI techniques. Moreover, none of the previous studies proposed a technical review for the use of this technology in CVD detection and prediction. Moreover, this review discusses in detail the challenges and future ideas in this field to help future researchers select the most relevant ideas on which to build their future work.

1.4.5. Key Findings and Contributions

Consequently, this article is rich in various new points that contribute to the body of knowledge on Multimodal ML:

- Discuss fusion and its fundamental role in defining the structure of Multimodal ML;
- Establish clear and precise boundaries to distinguish between Multimodal ML, traditional ML, multimodal datasets, multilabel models, and ensemble learning;
- Propose a new description for the different workflows that can be followed in the implementation of Multimodal ML algorithms;
- Discuss existing frameworks in the area of Multimodal ML and evaluate the contributions to this area;
- Review and discuss the state of the art of Multimodal ML in the diagnosis of CVDs;
- Examine the technical details associated with these implementations;
- Present completely and in detail the challenges that hinder Multimodal ML and the possible future perspectives that can be pursued to increase the efficiency of the technology.

2. Materials and Methods: What Is Multimodal ML?

The human mind processes information from multiple senses simultaneously. Sometimes it is not enough to just hear about a problem; individuals need to see it for themselves in order to make an informed judgment. For Artificial Intelligence to expand its knowledge of the world, it must be able to process a variety of information sources that may contradict each other. This principle also applies to the field of AI known as Machine Learning (ML), where Multimodal Machine Learning focuses on using numerous data sources to achieve a single goal by leveraging complementary information in a unified computational framework. The ability to explore diverse data increases predictive power and leads to more accurate and reliable results, making Multimodal Machine Learning a multidisciplinary topic with tremendous efficiency and amazing potential [5,10].

2.1. Overview and Definition(s)

Despite the fact that Multimodal Machine Learning is a popular and young research area that has received much attention, it is still in its infancy [4–6,45]. As a result, there is no single and universally accepted definition. Nevertheless, all definitions lead to the same concept: the ability to analyze different data sets to reach a single conclusion. For example, the authors describe in [4] Multimodal ML as the ability to evaluate data from Multimodal datasets, identify a common phenomenon, and use complementary knowledge to learn a complex task. Multimodal datasets are described in this way as data seen with many sensors, where the output of each sensor is called a modality and can be associated with a dataset. Similarly, the authors of [5] describe Multimodal ML as the integration of multiple data sources collected by different instruments, devices, or techniques, followed by the analysis of these merged data using different ML architectures. In addition, Multimodal Machine Learning is described in [10] as an area that aims to develop intelligent models that can process and link data from many sources.

2.2. Multimodal ML and Data Fusion

Multimodal ML brings together data from multiple and disparate modalities to identify a single task. The discipline behind merging data from multiple sources is called data fusion. More specifically, data fusion is defined as “the process of combining data to refine state estimates and predictions” [5]. According to the Joint Directors of Laboratories Data Fusion Subpanel (JDL), the technique of “data fusion” is a must for processing more than one type of data [46]. The authors in [46] support this definition by explaining that any process that deals with associating, correlating, or combining data from one or more sources to obtain enriched information is called a process that uses data fusion. In data fusion, given the novelty of the literature, there is no consensus on how best to combine different data, especially since there are four different techniques for implementing data fusion, which may have many names depending on the context and research area [5,46,47]. These different approaches are illustrated in Figure 3:

- Early Fusion: also called Low-Level Fusion, is the simplest form of data fusion in which disparate data sources are merged into a single feature vector before being used by a single Machine Learning algorithm. Therefore, it can be referred to as a multiple-data, single-algorithm technique.
- Intermediate Fusion: is also referred to as Medium-Level Fusion, joint fusion, or Feature-Level Fusion, and occurs in the intermediate phase between the input and output of a ML architecture when all data sources have the same representation format. In this phase, features are combined to perform various tasks such as feature selection, decision-making, or predictions based on historical data.
- Late Fusion: also known as decision-level fusion, defines the aggregation of decisions from multiple ML algorithms, each trained with different data sources. In addition, various rules are used to determine how decisions from different classifiers are combined, e.g.,:
  - Max-fusion
  - Averaged-fusion
  - Bayes' rule-based
  - Even rules learned using a metaclassifier
- Hybrid Fusion: defines the use of more than one fusion discipline in a single deep algorithm.

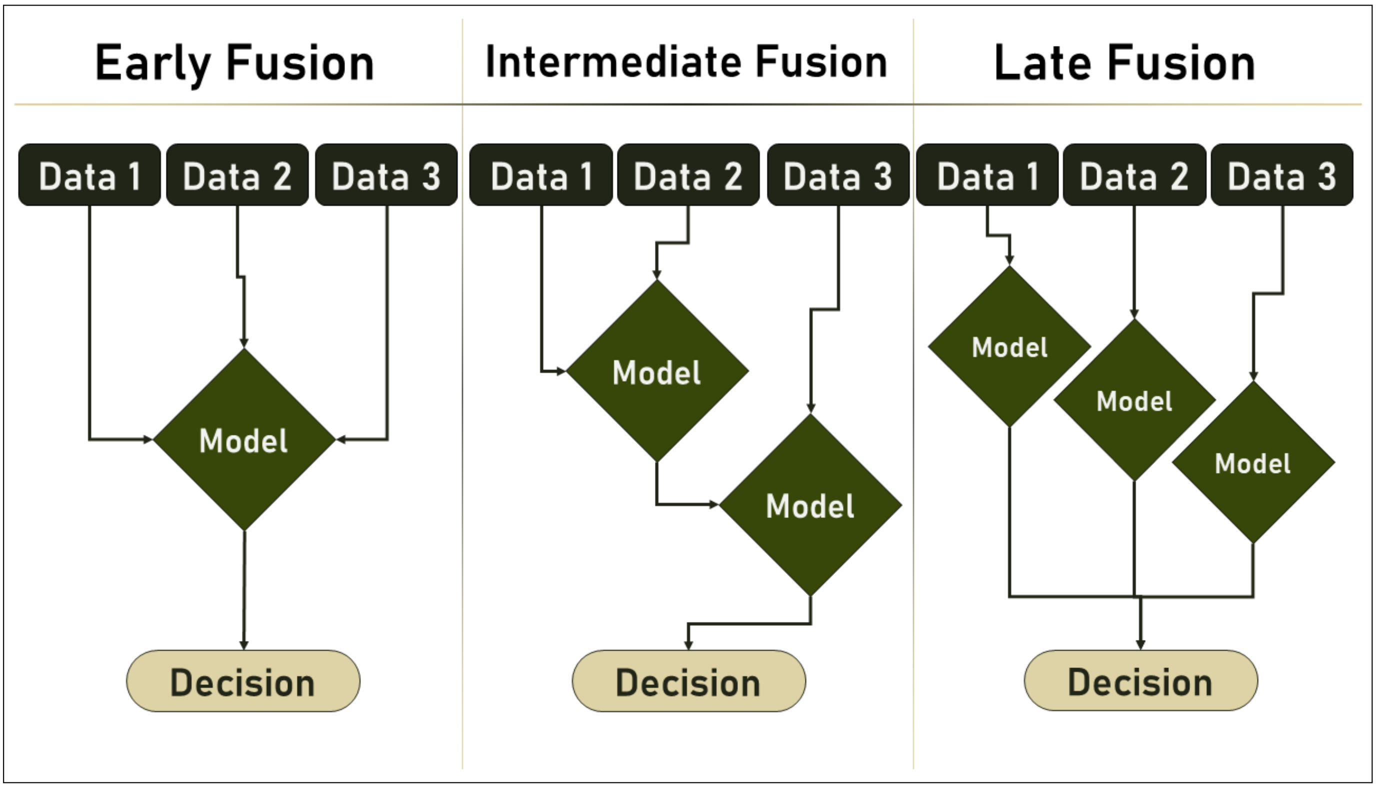


Figure 3. Data fusion approaches.

Based on the information in [4,5], early fusion is the most common form of fusion, which has the advantage of converting all data into the same representation that can be classified using robust classical models, such as Support Vector Machines or Logistic Regression. However, when the input modalities are particularly uncorrelated and have widely varying dimensionality and sampling rate, it is easier to use a late fusion approach. In addition, both early and late fusion offer the most flexibility in terms of the number of models that can be used to analyze the data, but there is no conclusive evidence that late fusion is better than early fusion because its performance is highly problem dependent. Alternatively, intermediate fusion provides more flexibility in terms of how and when representations learned from Multimodal data are fused. Table 2 discusses the different features of each approach.

Table 2. Data fusion approach specifications.

Attribute	Early	Intermediate/Joint	Late/Decision
Ability to handle missing data	no	no	yes
Scalable	no	yes	yes
Multiple models needed	no	yes	yes
Improved accuracy	yes	yes	yes
Voting/weighting issues	no	yes	yes
Interaction effects across sources	yes	yes	no
Interpretable	yes	no	no
Implemented in health	yes	yes	yes

2.3. Multimodal ML: Technical Perspectives

The goal of Multimodal Machine Learning, also known as Multimodal Deep Learning, is to develop algorithms and models that can interpret and learn from data across multiple modalities, such as text, audio, images, and video. Multimodal ML is a thriving research area with the potential to transform a wide range of applications, from speech recognition and language translation to autonomous cars and medical imaging, among many other areas. Multimodal ML, from a technical perspective, encompasses the various approaches, algorithms, and architectures used in creating and evaluating these models. Data preprocessing, feature extraction, model architecture, training methods, evaluation criteria, generalization, interpretability, and scalability are the most common possible viewpoints. Understanding the technical aspects of Multimodal ML is essential for developing efficient models that can leverage complementary instances across many modalities and make more accurate and robust predictions in the real world. Therefore, the technical perspectives of Multimodal ML are described below.

2.3.1. Data Preparation

Because Multimodal data are often complex and heterogeneous, they must be thoroughly processed before they can be used to train the model. The first step is to recognize the many modalities in action, then learn how to preprocess them, and finally, merge them into a single representation that can be fed into the model [4,5,10].

2.3.2. Model Architecture

Multimodal data can be represented in a variety of ways, including concatenation, fusion, and attentional mechanisms. Choosing the right architecture that can handle the multiple modalities and learn a combined representation is crucial depending on the data and the task to be solved [46,47].

2.3.3. Training Strategies

Pretraining individual modalities, joint training of all modalities, and training individual models and combining them at the time of inference are all viable options for training Multimodal ML models. Selecting the right training methods is a crucial step in achieving the desired goal [4,5,10].

2.3.4. Evaluation Metrics

Following the performance metrics used to evaluate classical ML algorithms, accuracy, precision, recall, sensitivity, specificity, F1 score, and area under the curve (AUC) are just some of the measures that can be used to evaluate Multimodal ML algorithms. It is controversial whether these measures are useful or not when applied to Multimodal data.

As a result, the use of evaluation criteria that consider the success of each modality and the overall performance of the model is essential [21–23].

2.3.5. Generalization

Multimodal models are often trained on a specific collection of data and may not generalize well to new data. To assess how well the model can be generalized, it should be tested and validated with data that are very different from the training data [21–23].

2.3.6. Interpretability

Because of their complexity and the relationships between multiple modalities, Multimodal ML algorithms can be difficult to understand and even more difficult to explain and interpret. To decipher the decision process of the model, some tools such as attentional mechanisms and visualization can be used [21–23,48].

2.3.7. Scalability

In Multimodal Machine Learning, scalability is critical because it enables models to deal with real-world situations where datasets are large and complex, and the amount of data is constantly growing. To ensure that the models can cope with the increase in data volume and complexity in the future, it is necessary to develop models that are scalable to enable effective training and deployment, reduce computational costs, and scale the models [25–27,48].

2.4. Multimodal ML and Other Technologies: Borderlines

Multimodal Machine Learning is a new and rapidly growing discipline that focuses on building models that can learn from a variety of data sources. To distinguish Multimodal ML from other areas of Machine Learning, its characteristic aspects should be highlighted, such as the use of many modalities and the need for effective integration of these modalities. Establishing precise terminology and creating an understandable description of the field will help to differentiate it from other techniques. However, because it is a relatively new field, there may be an overlap with other areas of Machine Learning, and it will be critical to accurately define the boundaries of Multimodal ML as the topic evolves.

2.4.1. Multimodal ML vs. Multimodal Datasets

Multimodal datasets are datasets acquired with different sensors, instruments, technologies, or devices to observe a common phenomenon, where the acquired data are considered complementary [49]. Consequently, multimodal datasets define the data itself, regardless of the identity of the algorithms used to analyze the data, whether they have a multimodal or unimodal architecture. However, merging multimodal datasets and unifying their representation into a single vector and then analyzing them with an ML model is considered an early fusion approach that is a type of Multimodal ML.

2.4.2. Multimodal ML vs. Multilabel Models

Multilabel Machine Learning algorithms are used to analyze datasets with more than one target variable. For example, the output of multilabel classification models consists of multiple classification labels. Moreover, when performing predictions using multilabel ML algorithms, a given input may belong to more than one label. For example, predicting the category of a movie may result in horror, action, science fiction, drama, or some or all of these categories simultaneously. In other words, multilabel classification associates data with a set of labels. Classification involves learning from a set of examples associated with a single label called “l” from a set of disjoint labels called “L”, where $|L| > 1$. When $|L| = 2$, the learning problem is called a binary classification problem, and when $|L| > 2$, it is called a multiclass classification problem [50,51]. Thus, Multimodal ML and multilabel learning differ in the data structure itself, where the former uses data from multiple or different sources to obtain a single result, while the latter uses data from only one source to obtain a single classification result with more than two possible outcomes.

2.4.3. Multimodal ML vs. Ensemble Learning

The goal of ensemble Machine Learning is to improve performance and accuracy by combining numerous models into a single prediction. When making predictions, ensemble learning uses multiple interconnected models rather than a single model. Ensemble learning combines the predictions of many models with the goal that the combined predictions are more accurate and robust than any single model. There are several types of ensemble learning techniques, including [52,53]:

- Bagging (Bootstrap Aggregating): is the process of training several models using random subsets of the training data to minimize overfitting;
- Boosting is a technique in which models are trained progressively, and the weights of misclassified data points are raised to enhance performance;
- Stacking is the process of training many models and combining their predictions with another model to obtain the final forecast.

Ensemble Learning has proven useful in a variety of applications, including classification, regression, and anomaly detection. Following this, although Ensemble Learning uses multiple ML models to solve one task, the main difference between these two technologies is that Multimodal ML is able to analyze more than one dataset with more than one model to solve a task, while Ensemble Learning uses multiple models for the same dataset to solve a task. Therefore, unlike Multimodal ML, Ensemble Learning does not perform data fusion to solve the task. Table 3 below summarizes the comparison between Multimodal ML and other technologies.

Table 3. Multimodal ML vs. other technologies.

Technology \Specs	Definition	Main Goal	Perform Better than ML	Merge Datasources	Merge Models
Multimodal Datasets	Datasets that include multiple modalities of data	Enable Multimodal Machine Learning	Not Applicable	Yes	Not Applicable
Multilabel Learning	A supervised learning technique in which an instance can be assigned to multiple labels	Accurately label instances with multiple labels	Not Applicable	No	No
Ensemble Learning	Combines multiple models to improve the accuracy of the prediction	Improve prediction Accuracy	Yes	No	Yes
Multimodal ML	Combines multiple types of models/data to improve performance and feasibility	Improve Performance	Yes	Yes	Yes

2.5. Multimodal ML Available Frameworks

Multimodal Machine Learning frameworks provide a systematic approach for developing models that can learn and integrate information from multiple modalities such as text, audio, images, and other data types. As more and more data are created across multiple modalities, multimodal frameworks for Machine Learning are becoming increasingly important. These frameworks enable the integration of diverse information, allowing for a more comprehensive understanding of complicated events. They’re used in everything from speech recognition and natural language processing to image and video analysis. Some of the existing and commonly used Multimodal ML frameworks are:

- MMF (a framework for multimodal AI models) [54]: is a PyTorch-based modular framework. MMF comes with cutting-edge vision and language pretrained models, a slew of ready-to-use standard datasets, common layers and model components, and training and inference utilities. MMF is also utilized by various Facebook product teams for multimodal understanding of use cases, allowing them to swiftly put research into production;
- TinyM^{2}Net (a flexible system, algorithm co-designed multimodal learning framework for tiny devices) [55]: a unique multimodal learning framework that can handle multimodal inputs of images and audio and can be re-configured for individual application needs. TinyM2Net also enables the system and algorithm to incorporate fresh sensor data that are tailored to a variety of real-world settings. The suggested framework is built on a convolutional neural network, which has previously been recognized as one of the most promising methodologies for audio and visual data classification;
- A Unified Deep Learning Framework for Multimodal Multi-Dimensional Data [56]: is a framework capable of bridging the gap between data sufficiency/readiness and model availability/capability. For successful deployments, the framework is verified on multimodal, multi-dimensional data sets. The suggested architecture, which serves as a foundation, may be developed to solve a broad range of data science challenges utilizing Deep Learning;
- HAIM (unified Holistic AI in Medicine) [57]: It is a framework for developing and testing AI systems that make use of multimodal inputs. It employs generalizable data preprocessing and Machine Learning modeling steps that are easily adaptable for study and application in healthcare settings.
- ML4VocalDelivery [58]: a novel Multimodal Machine Learning technique that uses pairwise comparisons and a multimodal orthogonal fusing algorithm to create large-scale objective assessment findings of teacher voice delivery in terms of fluency and passion;
- Specific Knowledge Oriented Graph (SKOG) [59]: a technique for addressing multimodal analytics within a single data processing approach in order to obtain a streamlined architecture that can fully use the potential of Big Data infrastructures' parallel processing.

2.6. Training and Evaluation of Multimodal ML Algorithms

Multimodal Machine Learning is a technique that combines different modalities in an attempt to solve a complex task. Given that Multimodal ML is based on the concept of data fusion [46], the training process of a multimodal model may differ depending on the type of fusion (early, intermediate, or late fusion). Although it is a Machine Learning concept, it follows the familiar ML workflow, which would be: data preprocessing, model selection, model training, evaluation, fine-tuning, and deployment, but different steps may occur depending on the fusion stage.

First, in the case of early fusion, after preprocessing, the different datasets can be combined and merged into one modality. Once the data are ready and fused, it can be fed into the model to be trained, and then the other steps can be performed. In the second case, called intermediate fusion, the data passed to the same model are merged after preprocessing, then a single model is trained on the fused dataset, and later, the result of the refined model is fused with other models if they exist. Finally, in the late fusion approach, each dataset is passed to a different model after preprocessing, then the models are trained, evaluated, and fine-tuned, and later, the results are merged into a single result. The three approaches are shown in Figure 4 below.

On the other hand, the evaluation of the Multimodal ML model is also influenced by the chosen approach of data fusion. Since data fusion applies a single model to fused data sources, only a single evaluation is required. In the other two approaches, intermediate and late fusion, each individual model must be evaluated, and later, the final model that merges the different models must be evaluated. The performance measures used to evaluate the Multimodal ML correspond to parameters commonly used in the classical ML domain, such as accuracy, precision, recall, sensitivity, specificity, F1 Score, Area Under Curve (AUC) and others [44]. The evaluation step is also shown in Figure 4 below.

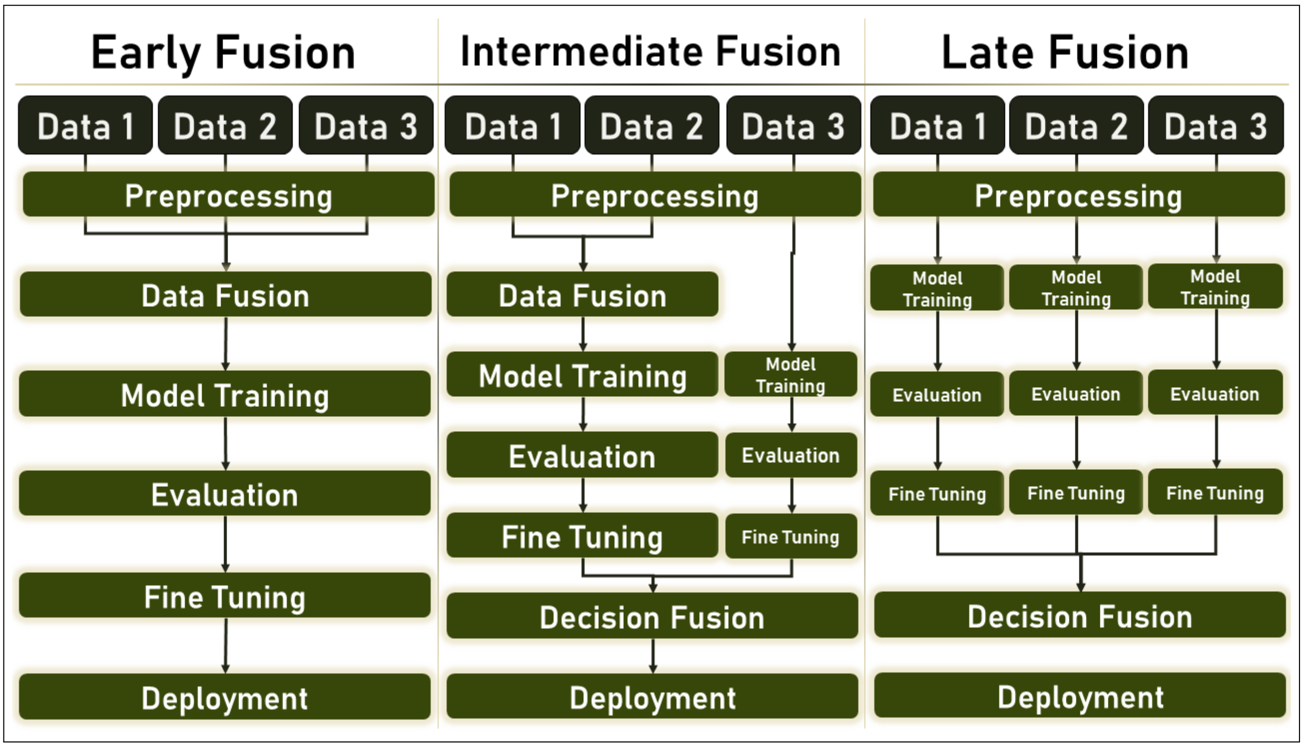


Figure 4. ML workflows based on Multimodal ML approaches.

3. Results: Multimodal ML in Action

Multimodal Machine Learning is a rapidly growing research area that involves the use of many modalities to evaluate and interpret complicated data, such as images, audio, and text [5,47]. Numerous real-world applications, including self-driving vehicles, voice recognition software, and medical imaging, require the ability to integrate and analyze data from multiple sources. Multimodal ML is based on the notion that multiple modalities provide complementary information and that merging these modalities can lead to more accurate and robust models. Multimodal ML has been a hot topic in the scientific community in recent years, and researchers have been striving to develop new algorithms and strategies to improve its performance [5,60–62].

3.1. Multimodal ML: Fields of Implementation

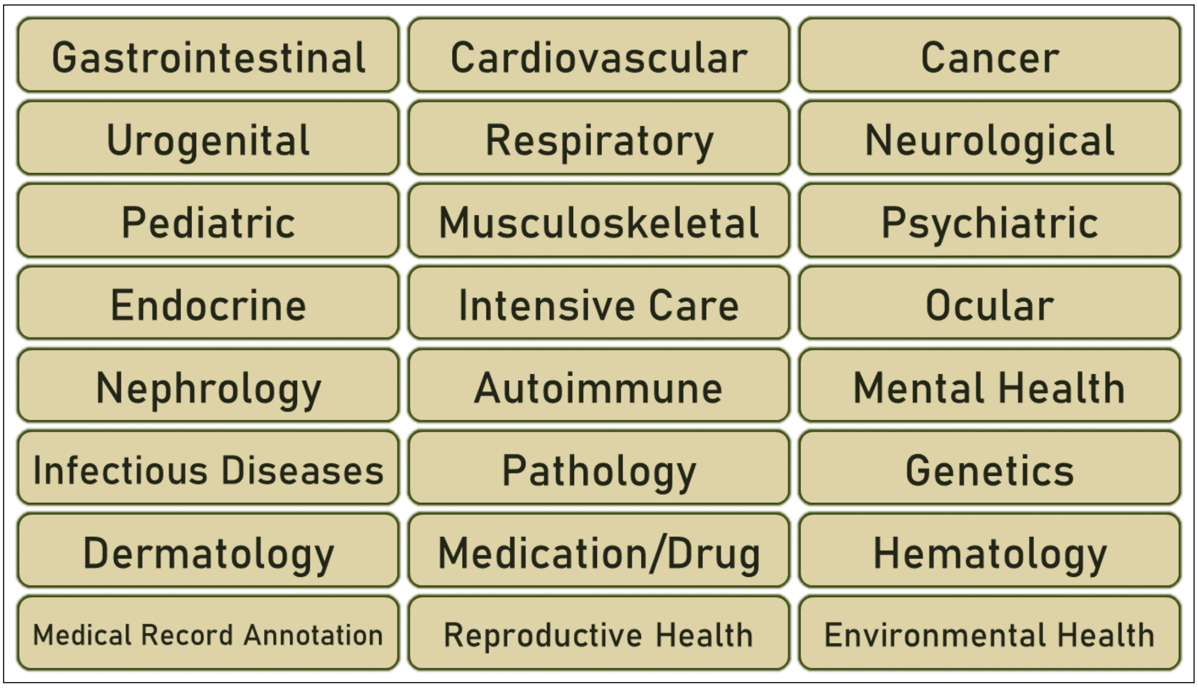
The ability to analyze diverse and complementary data increases the success of Machine Learning algorithms in solving more complex problems. In this context, Multimodal ML has proven its success in a variety of domains. Some of the most promising application areas include [5,60–64]:

- Healthcare: in medical imaging, Multimodal ML can be used to integrate information from different imaging modalities such as MRI, CT, and PET scans to improve diagnosis and treatment planning. It can also be used to classify and predict disease based on a mix of clinical, genetic, and imaging data;
- Autonomous Vehicles: by combining data from numerous sensors, the Multimodal ML can help self-driving vehicles better understand their surroundings. This has the potential to improve object recognition, navigation and safety;
- Natural Language Processing: by blending audio and text data, Multimodal ML can improve speech recognition and natural language comprehension. This can help voice assistants, chatbots and other applications improve their accuracy;

- Robotics: by combining inputs from sensors such as cameras, microphones, and touch sensors, Multimodal ML can be used to improve robot perception and interaction. This has the potential to improve navigation, object recognition, and human–robot interaction;
- Education: this technology is used in education to analyze student data from numerous sources, such as exams, quizzes, and essays, to make individualized learning suggestions and improve student performance;
- Agriculture: this technology can revolutionize agriculture by enabling the optimization of farming processes. It can be used for crop yield prediction, pest and disease detection, precision agriculture, and crop optimization by combining data from multiple sources, such as satellite imagery, weather data, and soil moisture sensors;
- Internet of Things (IoTs): this technology can be used in the context of the Internet of Things to make better use of data provided by networked devices. Multimodal ML can enable more accurate and robust models for predicting, monitoring, and managing IoT systems by incorporating data from many sources, such as sensors, cameras, and audio recordings, leading to advances in areas such as energy management, transportation, and smart cities.

3.2. Multimodal ML in Healthcare

Multimodal ML is still in its infancy but has been studied and applied in many areas of life, including healthcare. Multimodal ML is an effective method for assessing health data from multiple sources and improving predictive ability due to the inherent heterogeneity of such information [5,62,64]. To date, there are 128 applications of Multimodal ML in healthcare, with neurology and cancer being the most prevalent, as reported in [5]. Multimodal machine learning has shown promising results in various medical areas, as illustrated in Figure 5. While the areas depicted in the figure are the most commonly studied to date, it is worth noting that the potential applications of multimodal machine learning extend beyond these domains:



Gastrointestinal	Cardiovascular	Cancer
Urogenital	Respiratory	Neurological
Pediatric	Musculoskeletal	Psychiatric
Endocrine	Intensive Care	Ocular
Nephrology	Autoimmune	Mental Health
Infectious Diseases	Pathology	Genetics
Dermatology	Medication/Drug	Hematology
Medical Record Annotation	Reproductive Health	Environmental Health

Figure 5. Healthcare sectors where Multimodal ML has been implemented so far.

3.3. Multimodal ML and Cardiovascular Diseases: State-of-the-Art

Cardiovascular Disease, the most deadly disease, is a topic of interest for Multimodal ML implementations. For example, in [65], the authors developed a multimodal data fusion ML model to predict hypertension. Using a Convolutional Neural Network (CNN)-based model, they analyzed different Electronic Health Records (EHRs) that were merged with the multimodal data fusion approach. Their model proved its efficiency with an accuracy that reached 94%. In a similar approach, the authors in [66] created a multimodal data fusion model to predict 30-day hospital readmission of patients with heart failure. For this purpose, they developed a Deep Unified Network (DUNs) trained with EHRs from the Enterprise Data Warehouse (EDW) and the Research Patient Data Repository (RPDR). Their model achieved an accuracy of 76.4%. In addition, the study [67] also implemented a data fusion model to cluster patients with hypertension. The authors proposed a novel Hybrid Non-Negative Matrix Factorization (HNMF) method-based model trained with phenotype and genotype information from the HyperGen dataset [68]. The accuracy of their proposed model reached up to 96%. In addition, the authors also developed a data fusion model in [69]. Their goal was to classify different CVDs, so they developed and trained a Text–Image Embedding network (TieNet) model with Chest X-Ray and free-text radiology clinical reports extracted from ChestX-Ray14 [70] and OpenI [71] Chest X-Rays datasets. The proposed model had an Area Under Curve (AUC) of 0.9, as they mentioned. In the same context, the solution proposed in [72] is a data fusion model developed to classify patients at potential cardiovascular risk. The model was based on Recurrent Neural Networks and trained on EHR data, achieving 96% accuracy.

Other implementations proposed model fusion or hybrid multimodal ML architectures to solve their problems. For example, in [73], the authors proposed a hybrid fusion multimodal ML to predict various cardiac diseases such as atelectasis, pleural effusion, cardiomegaly and edema. They created several ML models to analyze radiographs and associated reports obtained from MIMIC-CXR [74] and OpenI [71] Chest X-Ray datasets. Their proposed solution proved to be better than old implementations in terms of accuracy. Similarly, in [75], a multimodal unsupervised learning approach was proposed for Cardiometabolic Syndrome Detection. The authors applied multimodal hybrid fusion by combining unsupervised ML models to analyze fused data from metabolome, microbiome, genetics, and advanced imaging. Furthermore, in [76], the authors proposed a multimodal fusion-based ML model for stroke prediction. They fused both 3D Convolutional Neural Network and Multilayer Perceptron models to analyze neuroimaging information and clinical metadata extracted from the Hotter [77] dataset, which proved to be efficient and powerful with an AUC of 0.90. In addition, the solution proposed in [78] was used to predict Pulmonary Embolism (PE) by fusing multiple ML models trained with Computed Tomography Pulmonary Angiography scans and EHRs. Their model recorded an AUC of 0.947. Furthermore, in [79], the authors developed a Recurrent Neural Network model with Bidirectional Long-Term Memory (BiLSTM) to predict cardiovascular risk. Their model was trained with EHR data extracted from the Second Manifestations of ARTerial Disease (SMART) Study [80] and recorded an AUC of 0.847.

Similarly, in [81], the authors developed a data fusion model to predict Acute Ischemic Stroke. They used a series of cardiac CT images with EHR recordings to train a Gradient Boosting classifier that achieved an AUC of 0.856. Similarly, the study [82] proposed a Deep Convolutional Neural Network (DCNN) data fusion model to analyze Electrocardiograph (ECG) and Chest X-Ray images to efficiently predict Accessory Pathways (APs) syndrome. Finally, in [83], the authors proposed a novel tensor-based dimensionality reduction method using Naive Bayes, SVM, Random Forest, Adaboost, and LUCCK models. The created models were trained with fused data composed of Salient Physiological Signals and EHR data. Their solution was able to predict Hemodynamic Decompensation with an AUC value of 0.89. Table 4 below summarizes and presents the Multimodal ML implementations in CVDs.

Table 4. Multimodal ML implementations in Cardiovascular Disease diagnosis and prediction.

Ref	Year	Type	Parameter Studied	Predicted Outcome	Model	Architecture	Datasets Used	Performance
[65]	2017	Classification	EHR Data	Hypertension	Convolutional Neural Network	Data Fusion	Private Data	Accuracy: 94.8%
[66]	2018	Classification	EHR Data	Thirty-day readmission risk for heart failure patients	Deep Unified Networks (DUNs)	Data Fusion	Enterprise Data Warehouse (EDW) Research Patient Data Repository (RPDR)	Accuracy: 76.4%
[67]	2018	Clustering	Phenotype and Genotype Information	Hypertension	Hybrid Non-Negative Matrix Factorization (HNMF) model	Data Fusion	HyperGEN dataset [68]	Accuracy: 96%
[69]	2018	Classification	Chest X-Ray Clinical Free-Text Radiological Report Scan	Several CVDs	Text-Image Embedding network (TieNet)	Data Fusion	ChestX-Ray14 dataset [70] OpenI Chest X-Ray dataset [71]	AUC: 0.9
[72]	2019	Classification	EHR Data	Cardiovacsular Risk Prediction	Recurrent Convolutional Neural Network	Data Fusion	obtained from a grade-A hospital of second class in Wuhan	Accuracy: 96%
[73]	2020	Classification	MIMIC-CXR Radiographs and Associated Reports	Atelectasis, Pleural Effusion, Cardiomegaly, Edema	four pre-trained Vision+Language models: LXMERT / VisualBERT / UNIER / PixelBERT	Hybrid Fusion	MIMIC-CXR Chest X-Ray Dataset [74] OpenI Chest X-Ray Dataset [71]	Enhanced accuracy of classification
[75]	2020	Clustering	Metabolome Microbiome Genetics Advanced Imaging	Cardiometabolic Syndrome	Combianation of unsupervised ML Models	Hybrid Fusion	Private Data	-
[76]	2020	Classification	Neuroimaging Information Clinical Metadata	Stroke	3D Convolutional Neural Network Multilayer Perceptron	Model Fusion	Hotter Dataset [77]	AUC: 0.90
[78]	2020	Classification	Computed Tomography Pulmonary Angiography Scans EHR	Pulmonary Embolism (PE)	Combination of ML Models	Hybrid Fusion	Data obtained from Stanford University Medical Center (SUMC)	AUC: 0.947
[79]	2020	Classification	EHR Data	Cardiovascular Risk	Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network	Hybrid Fusion	Second Manifestations of ARTerial Disease (SMART) Study [80]	AUC: 0.847
[81]	2020	Classification	Different Cardiac CT Images and EHR Data	Acute Ischemic Stroke	Gradient Boosting Classifiers	Data Fusion	obtained from Department of Neuroradiology at Heidelberg University Hospital (Heidelberg, Germany)	AUC: 0.856
[82]	2021	Classification	Electrocardiograph (ECG) Chest X-Ray	Cardiac Accessory Pathways (APs) Syndrome	Deep Convolutional Neural Network (DCNN)	Data Fusion	Private Data	-
[83]	2021	Classification	Salient Physiological Signals EHR Data	Hemodynamic Decompensation	Used a novel tensor-based dimensionality reduction with the below models: Naive Bayes SVM Random Forest Adaboost LUCCK	Data Fusion	Collected retrospectively from Michigan Medicine data systems	AUC: 0.89

3.4. Multimodal ML and CVDs: Discussion

Multimodal ML is a method for training different modalities using heterogeneous data that may not fit the same structure, format, or type that can be used for traditional ML algorithms. In the field of disease diagnosis, Multimodal ML could be used to train models on a huge distributed dataset of patient data from different hospitals or clinics. This method allows information and knowledge to be fused to solve complex problems. Using a larger, more diverse dataset also allows for more accurate and robust models. However, the implementation of Multimodal Machine Learning for disease prediction, especially Cardiovascular Disease, can be discussed from different angles, which are detailed in this section.

3.4.1. Models Performance: Competition between Multimodal and Classical ML

Data collection is the starting point for the operation of the established pipeline in the classical ML. It is generally accepted that more data can be used to increase the accuracy of an already trained Machine Learning model. It is generally accepted that due to the ability of Multimodal ML to analyze heterogeneous data, the accuracy of the models far exceeds that of typical ML models where more data are analyzed simultaneously.

In this context, the results presented in Table 4 reflect the high feasibility and accuracy that Multimodal ML cope with the diagnosis and prediction of Cardiovascular Disease. For example, the studies [65,67,72] achieved high accuracy records, with the first recording 94.8% and the other two, 96%. These results are highly comparable to the state of the art of conventional ML models used for the detection and prediction of CVDs and cerebrovascular events, with the highest recorded accuracy reaching 91.80%, as shown in [84]. In addition, the studies [69,76,78] recorded high values for Area Under Curve (AUC), with the first and second reaching a value of 0.9 and the third up to 0.95 for this parameter. These values demonstrate the high feasibility of these studies, which are consistent with and even exceed conventional ML algorithms. Moreover, the authors mention in [73] that their results show improved classification accuracy compared to conventional ML algorithms.

On the other hand, the results in [66] failed to outperform or even match conventional ML algorithms, where the recorded accuracy was 76.4%, which is lower than the values obtained by the latest ML algorithms in predicting ML models [84]. In addition, the studies [79,81,83] obtained different AUC values of 0.85, 0.86, and 0.89, respectively. These values are high and feasible, but they are close to but do not exceed the highest results obtained with classical ML models. Finally, the studies [75,82] did not mention the results obtained, which makes it impossible to compare their results with the classical ML models in the field of CVD diagnosis.

Overall, of the thirteen studies presented in Table 4, seven exceeded the results of the classical ML in terms of accuracy, three matched those results, and only one was obviously lower than them, and the other two are not comparable because they did not report their results. In this context, these figures help to confirm the hypothesis that the ability to analyze heterogeneous data increases the performance and accuracy of the models, which is a major strength in the field of multimodal ML since more than three-quarters of the Multimodal ML algorithms either match or exceed the results of the classic ML in the diagnosis of Cardiovascular Disease.

3.4.2. Real World vs. Research Implementations

The concept of Multimodal ML can be traced back to the early 2000s in the technology field, where authors in [85] suggested using this concept because the combination of communication modalities and acquisition devices can produce a wide range of unimodal and multimodal interface techniques. However, advances in computer technologies, data transmission, communication techniques, and other aspects have helped to increase the efficiency of Multimodal ML technology.

As a result, studies [65,75,85] have used their own data. Although these datasets are not publicly available, the authors assured that the data are real datasets collected from various health centers in compliance with medical standards and norms. This confirms that these studies can be classified as real-world studies. The same is true for [66,72,78,81,83], where each study used a dataset collected in different medical facilities in compliance with standard medical norms, making these studies real-world implementations.

On the other hand, the studies [67,69,73,76,79] used publicly available datasets, which are listed in Table 4. Although these datasets were collected under real-world conditions and obtained from patients, the study itself cannot be described as a real-world implementation. Real-world use of multimodal ML models in healthcare can provide a number of significant benefits, including:

- Improved Diagnostic Accuracy: Multimodal ML models can evaluate multiple sources of patient data, such as medical imaging, electronic health records, and genetic information, to make more accurate and thorough diagnoses. This can help physicians identify diseases and conditions at an early stage when they are more curable;
- Personalized Treatment: multimodal ML models can be trained on large data sets to identify trends and predict outcomes for individual patients. This can help physicians tailor treatments and therapies to the unique needs of each patient, leading to better outcomes and fewer side effects;
- Efficient Resource Allocation: Multimodal ML models can help physicians allocate resources more efficiently by identifying patients who are at higher risk for poor outcomes or need more intensive care. This has the potential to reduce healthcare costs while improving overall system efficiency;
- Improved patient experience: Multimodal ML models can help clinicians identify patients who need more individualized care or are at risk for problems or adverse events. This can help improve patient satisfaction and overall quality of care.

Overall, real-world adoption of Multimodal ML models in healthcare has the potential to enhance patient outcomes, lower costs, and improve healthcare delivery efficiency. However, it is critical that these models be created and used in an ethical manner, with proper protections for patient privacy and data security. That being said, the progress of Multimodal ML implementations and their real-world execution are promising where most of the carried applications are applied outside of labs, with real data, which enhances the trust in this technology and assists its adoption in the production stages.

3.4.3. Use of Smart Wearables and IoTs

Continuous monitoring of patients' heart rate, blood pressure and other biometric data through smart wearables and Internet of Things devices could revolutionize medical treatment. This has the potential to enable earlier detection of medical problems, more accurate diagnosis, and more personalized treatment approaches. Wearable technologies that can monitor and interact with the user's health could enable individuals to participate more fully in their treatment. In addition, Internet of Things (IoT) devices can enable physicians to monitor patients remotely and deliver treatments more effectively, reducing demand on healthcare systems and improving access to care for people in underserved or extremely remote and isolated areas. Smart wearables and Internet of Things (IoT) devices could increase hospital efficiency, save costs, and improve patient outcomes [86,87].

Consequently, only studies [67,75] considered the use of smart wearables or IoTs devices in their implementations. The other studies used data collected with other devices. Therefore, there is a lot of catching up to do in the implementation of multimodal ML in wearables and IoTs for CVD detection and prediction. Considering the fact that these technologies can revolutionize healthcare, as mentioned earlier, there is a great need to increase the use of wearables and IoTs in this field. In Table 5 below, the comparison between the performance of Multimodal ML and classical ML, the validation in practice, and the use of smart wearables and IoTs for the state of the art in predicting CVDs with Multimodal ML is summarized.

Table 5. Key findings in state-of-the-art of Multimodal ML in CVDs diagnosis.

Ref#	Multimodal ML Beats ML (Performance)	Real-World Implementation	Smart Wearables/IoTs Included
[65]	Yes	Yes	No
[66]	No	Yes	No
[67]	Yes	Public Dataset(s)	Yes
[69]	Yes	Public Dataset(s)	No
[72]	Yes	Yes	No
[73]	Yes	Public Dataset(s)	No
[75]	Not Available	Yes	Yes
[76]	Yes	Public Dataset(s)	No
[78]	Results Match	Yes	No
[79]	Results Match	Public Dataset(s)	No
[81]	Results Match	Yes	No
[82]	Not Available	Yes	No
[83]	Results Match	Yes	No

3.4.4. Limitations in the Use of Multimodal ML for Disease Prediction

From this perspective, the use of Multimodal Machine Learning for the diagnosis and prognosis of CVDs is still in its infancy. Apart from the fact that not all implementations of Multimodal Machine Learning are superior to traditional ML models, vivid real-world examples can be observed when discussing this topic. Moreover, it has been rare to see FL researchers using smart wearables or IoTs in their experiments. This highlights the need to further investigate the use of such technologies due to their high degree of practicality and applicability in the field. Other limitations and difficulties encountered in the field of multimodal ML and its applications in disease prediction are discussed in Section 4.1, which can also be seen below.

3.5. Multimodal ML in CVDs: A Technical Overview

In Multimodal Machine Learning technology, the main goal is to analyze different data with different structures, such as merging EHR data with medical images to predict the occurrence of Cardiovascular Disease. In this context, each Multimodal ML implementation follows its own workflow and goes through its own steps to achieve its goal. In the aforementioned implementations of Cardiovascular Disease detection using Multimodal ML, different workflows, model structures, and hyperparameters were used for different implementations. All the related data provided by the authors are listed in Table 6 below.

Table 6. Technical details for Multimodal models used in the prediction of CVDs.

Ref#	Model	Workflow Description	Training Parameters
[65]	CNN-Based Multimodal Disease Risk Prediction (CNN-MDRP) Algorithm	1. Data Representation: text is represented in the form of vector 2. Convolution Layer: perform convolution operation on vectors of 5 words 3. Pool Layer: use the max pooling (1-max pooling) operation on the input of the convolution layer 4. Full Connection Layer: pooling layer is connected with a fully connected neural network 5. Classifier: the full connection layer links to a softmax classifier	Iterations: 200 Sliding Window: 7 Running Time: 1637.2 s
[66]	Deep Unified Networks (DUNs)	1. All inner layers of DUNs can learn the prediction task from the training data to avoid over-fitting 2. The DUNs architecture has horizontally shallow and vertically deep layers to prevent gradient vanishing and explosion 3. There are only two horizontal layers from the data unit nodes to the output node, regardless of how many layers deep the architecture is vertically 4. Only the harmonizing and decision units have learning parameters	Number of epochs: 100 Number of inner layers: 5 Number of inner neurons: 759 Number of maxout: – Activation function: Sigmoid Dropout rate of: input layer: 0.397/inner layers 0.433

Table 6. *Cont.*

Ref#	Model	Workflow Description	Training Parameters
[67]	Hybrid Non-Negative Matrix Factorization (HNMF) model	1. Impute missing values in the phenotypic variables 2. For genetic variants, first annotate the variants and then keep those that are likely gene disruptive (LGD) 3. The preprocessed phenotypic measurements and genetic variants are then used as input to the HNMF model 4. The patient factor matrix is then used as the feature matrix to perform regression analysis to predict main cardiac mechanistic outcomes	Up to 50 iterations
[69]	Text–Image Embedding Network (TieNet)	1. Data Preprocessing and word embedding 2. Training TieNet model 3. Joint Learning for results fusion 4. Evaluation	Dropout: 0.5 L2 Regularization: 0.0001 for. Adam optimizer with a mini-batch size of 32 Learning Rate of: 0.001 Hidden Layer with 350 units
[72]	Recurrent Convolutional Neural Network	1. Structured Data: extract relevant data, supplement missing data, make correlation analysis to look for the relation among data and apply dimension reduction to obtain corresponding structured features 2. Unstructured Textual Data: first, use numerical values to present unstructured textual data based on work embedding. Then, the features of textual data are extracted based on RCNN 3. Use Deep Belief Network (DBN) to fuse features and predict disease risks	up to 200 iterations
[73]	VisualBERT, UNITER, LXMERT, and PixelBER	1. The feature map (7 × 7 × 1024) of CheXNet is first flattened by spatial dimensions (49 × 1024) then down-sampled to 36 1024-long visual features 2. Models are then trained with the data 3. Results are fused	Epochs: PixelBERT: 18 / other 3 models 6 SGD optimizer weight decay 5×10^{-4} learning rate 0.01 Each model can be fit into 1 Tesla K40 GPU when using a batch size of 16
[75]	Collection of unsupervised ML models	1. Data collection and data features 2. Data preprocessing 3. Network analysis 4. Key biomarker selection and Markov network construction 5. Stratifying individuals with similar biomarker signatures 6. Validation cohort	-
[76]	3D Convolutional Neural Network Multilayer Perceptron	All models were trained on a binary classification task using binary cross-entropy loss	Loss function: Binary cross-entropy loss Adam optimizer Initial weights were sampled from a Glorot uniform distribution Output layer activation function: Softmax function Early stopping used to prevent over-fitting
[78]	Different ML models	Seven different workflows based on the difference between models	Batch Size: 256 Epochs: 200
[79]	Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network	1. Embedding Layer: To extract the semantic information of radiology reports 2. Bidirectional-LSTM Layer: to achieve another representation of radiology reports 3. Dropout 4. Concatenation Layer 5. Dense Layers	Embedding dimension (d): 500 #neurons in LSTM layer: 100 CNN filter size: 5 filters in CNN: 128 neurons in dense layers: 64 Dropout rate: 0.2 Recurrent dropout rate: 0.2 Batch size: 64 epochs: 20 Optimization method ADAM
[81]	Gradient Boosting Classifiers	Integrative assessment of clinical, multimodal imaging, and angiographic characteristics with Machine Learning Allowed to accurately predict the clinical outcome following endovascular treatment for acute ischemic stroke	-
[82]	Deep Convolutional Neural Networks (DCNN)	First Model to analyze ECG 1. Convolutional Neural Network (CNN) 2. A one-dimensional CNN model was used to input the ECG data 3. The network model contained 16 convolution layers Followed by a fully connected layer 4. Then a Softmax layer, which calculated the probability of each of the four as the output in the last layer Second Model to analyze X-Ray images 1. A two-dimensional CNN model Then apply fusion to merge results	First Model Parameters: Adamax optimizer with the default parameters $\beta1$= 0.9, $\beta2$ = 0.999, and a mini-batch size of 32

Table 6. *Cont.*

Ref#	Model	Workflow Description	Training Parameters
[83]	Random Forest Naive Bayes Support Vector Machine Adaboost Learning Using Concave and Convex Kernels (LUCCK)	1. Apply feature extraction on fused data composed of Salient Physiological Signals and EHR data 2. Apply Tensor reduction functionality 3. Train the Machine Learning model	Naive Bayes: (NB) no hyperparameter tuning was trained Support Vector Machines: used linear, radial basis function (RBF), and 3rd-order polynomial kernels Random Forest: number of trees: 50, 75, and 100/minimum leaf size: 1, 5, 10, 15, and 20/node splitting criterion: cross entropy and Gini impurity/number of predictors to sample: [10, 20, . . . , 100]/maximum number of decision splits for the decision trees: 0.25, 0.50, 0.75, or 1.0 Adaboost: learning rate: 1

4. Discussion: Challenges and Future Perspectives

Recently, Multimodal Machine Learning (ML) has emerged as an effective method for studying and analyzing complex data from multiple sources and modalities. However, dealing with diverse data presents researchers with unique challenges that must be overcome for efficient analysis and interpretation to increase the feasibility and usability of multimodal ML [10,48,49,62]. Unifying and standardizing multiple data sources and establishing links between them are significant obstacles. In addition, data must be normalized and preprocessed to ensure reliability and accuracy. However, future research could take several approaches to mitigate these challenges and overcome future obstacles. This section addresses these issues and identifies future perspectives needed to overcome them and improve multimodal FL.

4.1. Challenges

Multimodal Machine Learning still struggles with various challenges arising from the use of heterogeneous data with different structures and formats. Moreover, the fusion process, whether applied to the data itself or to different trained models to recognize a single result, is a challenging process that requires further research. Therefore, the most common challenges can be summarized in the following points [10,48,49,62].

4.1.1. Data Availability and Quality

To efficiently train multimodal ML models, large amounts of high-quality data are needed. However, collecting and processing large amounts of high-quality data in healthcare can be challenging, especially for rare or complex diseases. Data scarcity or poor data quality can lead to biased or unreliable models, compromising the accuracy of predictions and treatment decisions. To develop more robust and effective multimodal ML models for healthcare, researchers must seek to identify and address data quality and quantity issues.

4.1.2. Data Representation

Multimodal ML promotes the use of data from multiple sources for presentation. As a result, there is a high likelihood of dealing with heterogeneous data, which presents a number of problems. For example, it may be difficult to merge heterogeneous data that do not overlap in common characteristics or overlap only in a very limited area. In addition, data from different sources may need to be processed to different extents, especially with respect to noise reduction and missing data management. This hurdle is clearly reflected in the fact that until recently, most multimodal representations were simply the concatenation of unimodal ones [88]. Smoothness, temporal and spatial coherence, sparsity, and natural grouping have been cited by authors in [89] as qualities for excellent data representation.

4.1.3. Data Integration and Interoperability

Multimodal Machine Learning models are used to integrate and analyze data from multiple sources, such as electronic health records, medical imaging, and genetic data.

However, data from different sources may use different formats, standards, or terminologies, posing significant challenges for data integration and interoperability. Medical images, for example, may use different file formats or imaging techniques, making it difficult to compare and analyze data from different studies or sources.

4.1.4. Fusion

It is not easy to learn the ability to merge information from two modalities and determine the optimal fusion strategy. This is due to the different predictive capacities and noise structures of the different information coming from different senses. In addition, the ability to deal with missing data at different levels has a significant impact on the ability to perform fusion tasks.

4.1.5. Translation

The challenge in translation is not only the heterogeneity of data but also the relationships between modalities. The translation or mapping of data is subjective; for example, two models may describe the same image in more than one correct way, and a perfect or uniform translation or mapping may not exist. Several studies argue that while translations can be quite broad and modality-specific, they still have a number of unifying features. Accordingly, there are two forms of translation, namely the "Example-Based" and the "Generative" models. The former relies on a dictionary to translate data across modalities, while the latter relies on the creation of a model that manages translation according to uniform or at least explicit standards.

4.1.6. Alignment

Finding connections and correspondences between subelements from two or more different modalities is called multimodal alignment. This also involves distinguishing between these linear connections rather than just recognizing them. In this context, there are few data sets with obvious and identifiable correlations. Therefore, it is challenging to perform similarity measurements across modalities. Moreover, there may be numerous alignments without being able to select the optimal one, and not all components in one modality may match in another.

4.1.7. Explainability and Interpretability

Multimodal Machine Learning models (ML) have shown great promise in healthcare by enabling more accurate and tailored diagnosis and treatment recommendations. However, these models can be very complicated and difficult to understand, making it difficult for physicians to understand how the models arrived at a particular decision or recommendation. The lack of interpretability and openness of these models can affect their clinical acceptance and confidence.

4.1.8. Co-Learning

Merging different modalities, such as images, text, and sensor data, can increase model performance and enable more comprehensive analysis of complicated data in Multimodal Machine Learning. However, there are significant hurdles to this fusion, including the difficulty of transferring knowledge, representation, and predictive models across modalities. Each modality has its own characteristics and advantages, and it can be difficult to successfully integrate these aspects into a coherent representation. In addition, different modalities may require different strategies for feature engineering, preprocessing, and modeling.

4.1.9. Increased Computation Cost

When multiple modalities and features are introduced into a Multimodal Machine Learning model, the complexity of the model may increase, and the performance of the model may degrade due to the increased difficulty in computing the desired outcome.

Complex models have higher processing requirements, which can increase inference times and memory consumption. The complexity of a model makes it more difficult to optimize, which can lead to an increased risk of over- or under-fitting the data.

4.1.10. Regulatory and Ethical Considerations

Apart from the technical hurdles in developing and implementing multimodal ML models in healthcare, there are also legal and ethical factors to consider. Depending on their intended use, these models may be subject to regulatory restrictions, such as the European Union's General Data Protection Regulation (GDPR) [90], China's Cyber Security Law of the People's Republic of China [91], the General Principles of the Civil Law of the People's Republic of China [92], the PDPA in Singapore [93], and hundreds of principles that apply around the world. In addition, researchers and clinicians must ensure that these models are created and used in an ethical manner and that patient privacy and data security are adequately protected. For example, patient data must be de-identified and protected from illegal access or disclosure. In addition, maintaining the fairness and openness of these models is critical to minimize bias and discrimination. Responsible development and adoption of multimodal ML models therefore require careful evaluation of these legal and ethical factors to ensure that they deliver safe, effective, and fair outcomes for patients.

4.1.11. Implementation and Adoption

To fully deliver on their promise to improve healthcare, Multimodal Machine Learning models (ML) must be integrated into current healthcare processes and systems. However, several barriers stand in the way of this integration, such as technological, organizational, and cultural. In addition to the technical challenges mentioned above, resistance to change, lack of stakeholder participation, and concerns about accountability and obligations are all examples of organizational and cultural hurdles that may arise.

These challenges give rise to the study questions in the list below (the abbreviation RQ in the list below refers to the term "research question"):

- **RQ1:** Multimodal ML needs sufficient data to be trained. Are the needed data sets available? And is their quality acceptable?
- **RQ2:** Multimodal ML deals with heterogeneous data that has different formats and structures. What approaches can be taken to represent the data used in this technology?
- **RQ3:** How can the heterogeneous data used in Multimodal ML be integrated and shared?
- **RQ4:** What are the best approaches for fusion, and how to choose between the different options available?
- **RQ5:** Given that different models can lead to the same result in different ways, how does one choose the optimal path?
- **RQ6:** How to align and link two different modalities, especially in the middle and late fusion cases?
- **RQ7:** The Multimodal ML is known for its black box identity. Is there a way to explain the methods by which a model arrives at its result?
- **RQ8:** In Multimodal ML, different models can be integrated to solve a complex task. What techniques can be applied to ensure efficient knowledge transfer between these models?
- **RQ9:** Heterogeneity and diversity in both models and data add to computational costs. How can this problem be dealt with to improve the usability and feasibility of the models?
- **RQ10:** How to ensure data exchange between multimodal ML facilities to comply with existing regulations and laws?
- **RQ11:** How can trust in multimodal ML be strengthened to promote its adoption in different areas of life?

4.2. Future Perspectives

The challenges faced in Multimodal Machine Learning can be solved through different approaches and perspectives. These solutions have either already been considered but should be more widely used in the field of Cardiovascular Disease prediction to improve and increase their usability and feasibility. In this context, the following solutions can serve as future recommendations.

4.2.1. Use Convenient Tools to Collect More Data

Modern technology has changed the method of data collection and analysis. The use of smart wearables and Internet-of-Things (IoT) devices has enabled the real-time collection of vast amounts of data [33,39,86,87]. These data can provide useful insights in a variety of areas, particularly in healthcare. In addition to these new data sources, current data sources should be used to create more complete databases. Researchers can gain access to larger and more diverse data sets by collaborating with other institutions, which can help them identify patterns and correlations that would not be obvious with smaller data sets. Collaboration between different institutions could be achieved using a variety of techniques such as Federated Machine Learning technology, which can help train Machine Learning models by sharing parameters rather than the data itself [9].

4.2.2. Automate and Boost Data Preprocessing

Creating larger and more comprehensive datasets could help improve the quality of Machine Learning models but is not yet sufficient. To gain valuable insights, data must be processed and analyzed using advanced techniques. These techniques include artifact automation and noise removal, as performed in [94,95]. In addition, it may be necessary to use techniques such as data augmentation [96] or data normalization [97] and data resampling [98] to ensure that the data are balanced and ready for model training and to improve the quality of the overall process.

4.2.3. Employment of Advanced Data Integration Tools

To address the problems posed by the diversity of data formats and structures, improved methods for data harmonization [99], standardization [100], and normalization [97] need to be developed, as well as the use of AI and ML algorithms to automate these processes. Multimodal ML has the potential to revolutionize healthcare by enabling thorough and tailored analysis of patient data from numerous sources if these barriers are overcome.

4.2.4. Embedding Modern Techniques to Enhance Explainability

To address the problems associated with the black-box nature of multimodal ML models, more explainable and interpretable models are needed that give healthcare professionals insight into how the models arrive at their judgments. Approaches such as feature relevance ranking [101], model visualization [102], decision rules [103], probabilistic [104] and neuro-fuzzy approaches [105], and many others can improve the interpretability of multimodal ML models so that interested parties can make more informed and confident treatment decisions. In the list below, a brief definition for each of these tools is presented:

- Feature relevance ranking: include methods such as permutation significance and partial dependency plots to give insights into the importance and correlations of input variables, allowing for a better understanding of the model's decision-making process and boosting transparency and interpretability in healthcare applications;
- Model visualization: such as decision trees and heatmaps that provide a graphical representation of the model's decision-making process, allowing for better understanding of the factors that influence the model's predictions and increasing the transparency and interpretability of the technology;

- Decision rules: by providing clear and understandable rationales for the model's predictions, decision rules that specify explicit decision criteria based on the input data improve the interpretability and transparency of machine learning models in healthcare.
- Probabilistic approach: employ probabilistic reasoning to represent and manage the uncertainty inherent in medical data allowing for transparent decision-making that can be easily understood by healthcare practitioners;
- Neuro-fuzzy techniques: combine the benefits of neural networks and fuzzy logic to generate more interpretable models that can deal with imprecise and uncertain inputs.

4.2.5. Implementing Necessary Methods to Guarantee Knowledge Transfer

The diversity of datasets and models in the field of multimodal ML can lead to knowledge transfer problems. Therefore, researchers need to develop novel strategies for multimodal feature selection [106], fusion [46], and modeling that can capture complementary information from many modalities while minimizing redundancy or overfitting. Overcoming these obstacles will allow for more robust and accurate multimodal ML models that will lead to improved diagnosis, treatment, and patient outcomes in healthcare settings.

4.2.6. Reducing Computation Cost

Reducing computational costs in multimodal ML is a critical issue. Therefore, researchers need to explore methods for model compression [107] and optimization [108] that can reduce the computational complexity of the model without compromising its performance. As an added bonus, Multimodal Machine Learning can benefit from efficient hardware and software implementations, such as specialized hardware accelerators and distributed computing frameworks, that can reduce computational load. The use of such techniques can help build multimodal ML models that are more robust, efficient, and scalable, and therefore applicable to a wider variety of health problems, leading to faster and more accurate solutions.

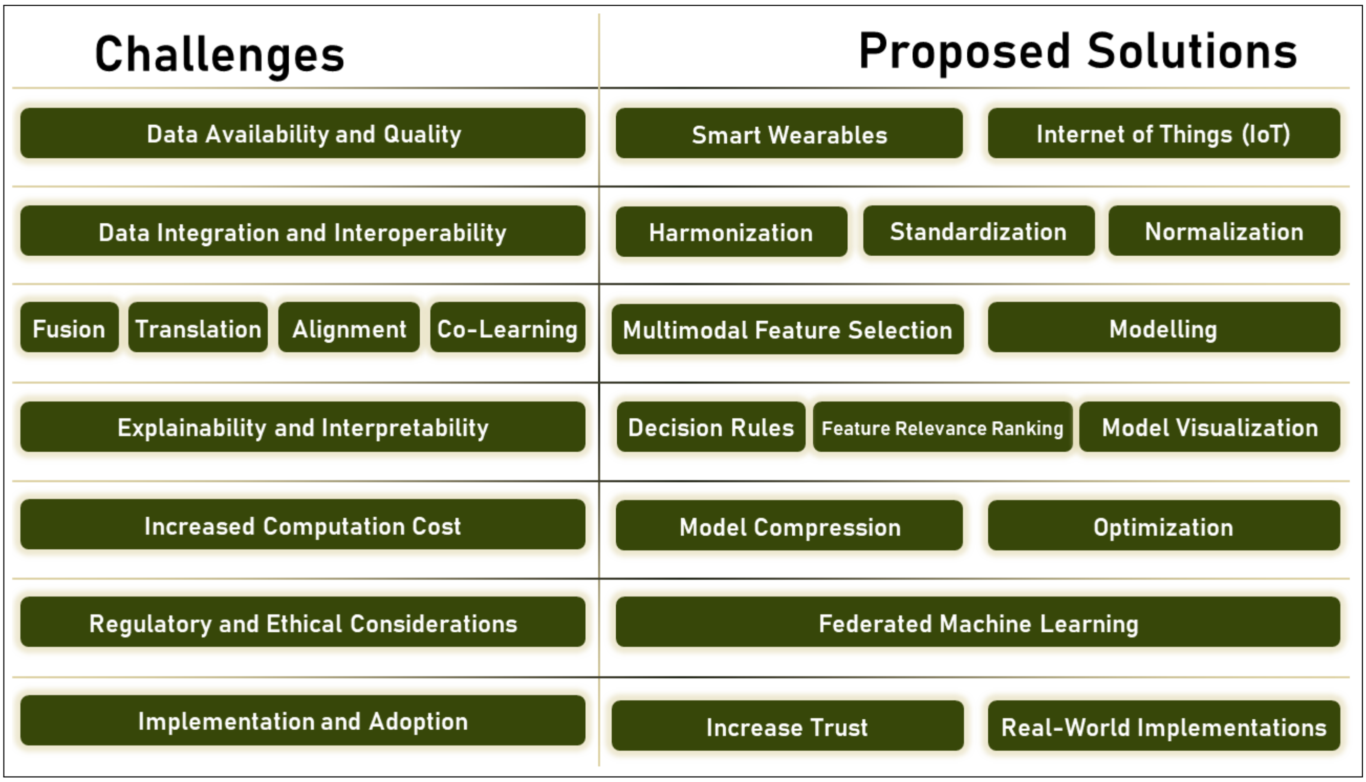
4.2.7. Increase Trust and Feasibility to Raise the Technology Adoption

Researchers, clinicians, information technology experts, and healthcare administrators must work together to increase confidence in multimodal ML technology. In addition, cultural and organizational barriers can be reduced by promoting trust and transparency through open dialog and training. The best way to improve patient outcomes and revolutionize healthcare delivery is to properly integrate multimodal ML models into current healthcare delivery processes and systems.

The results of the mapping of challenges and solutions can be summarized in the following topics (the symbol TR in the list below refers to the term "Trending Research Topic"):

- **TR1:** Data collection tools such as smart wearables and IoTs are very helpful in augmenting the data collected for multimodal ML algorithms;
- **TR2:** Data harmonization, standardization, and normalization are highly feasible for integrating heterogeneous data in the multimodal ML domain;
- **TR3:** Multimodal feature selection and modeling are techniques that can help ensure knowledge transfer between different modalities in a multimodal ML system;
- **TR4:** For better explainability and interpretability of a multimodal ML model, decision rules, feature relevance ranking, and model visualization are practical and feasible methods;
- **TR5:** Model compression and optimization are great tools for reducing computational costs in multimodal ML;
- **TR6:** Current and trending ML topics, such as Federated Machine Learning, can help overcome privacy and confidentiality issues in the Multimodal ML domain;
- **TR7:** Increasing feasibility, improving performance, and implementation in real-world scenarios are all factors that can help expand the adoption of multimodal ML technology in healthcare and, in particular, in Cardiovascular Disease detection.

Finally, the challenges that hinder the progress of Multimodal Machine Learning techniques, along with the solutions and future perspectives that could be pursued, are presented in Figure 6 below.



Challenges	Proposed Solutions
Data Availability and Quality	Smart Wearables Internet of Things (IoT)
Data Integration and Interoperability	Harmonization Standardization Normalization
Fusion Translation Alignment Co-Learning	Multimodal Feature Selection Modelling
Explainability and Interpretability	Decision Rules Feature Relevance Ranking Model Visualization
Increased Computation Cost	Model Compression Optimization
Regulatory and Ethical Considerations	Federated Machine Learning
Implementation and Adoption	Increase Trust Real-World Implementations

Figure 6. Multimodal machine learning challenges–solutions mapping.

5. Conclusions

In summary, Multimodal ML is a new technique that enables the simultaneous use of multiple models and data types in the creation of complex ML and DL models. Multimodal ML has the potential to significantly improve the accuracy and effectiveness of AI applications, especially in healthcare, where it has already become an important part of everyday patient care by addressing the problem of data heterogeneity. In particular, the technical features of Multimodal ML, such as data fusion and workflows, were covered, and the differences with other technologies, such as Ensemble Learning, were highlighted. In addition, an overview of the application of Multimodal ML in the diagnosis and prediction of Cardiovascular Disease was provided, highlighting the encouraging results to date and the room for growth in this area. Privacy, bias, and interpretability of results are just some of the remaining difficulties that need to be addressed, as with any rapidly evolving technology. However, it is likely that these obstacles can be addressed through further research and development and that multimodal ML will continue to play an important role in the development of AI applications in a variety of sectors, particularly healthcare.

Author Contributions: Conceptualization: M.M. and M.A.; formal analysis: M.M.; investigation: M.M.; methodology: M.M. and M.A.; supervision: M.A., A.B., H.I. and A.R.; visualization: M.M.; writing—original draft: M.M.; writing—review and editing: M.A., A.B., H.I. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 06351.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: We acknowledge the support of Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moor, J. The Dartmouth College artificial intelligence conference: The next fifty years. *AI Mag.* **2006**, *27*, 87.
2. Simone, N.; Ballatore, A. Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence* **2020**, *26*, 3–18.
3. John, M. *What Is Artificial Intelligence*; Stanford University: Stanford, CA, USA, 2007
4. Ramachandram, D.; Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [CrossRef]
5. Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Luo, Y. Multimodal Machine Learning in Precision Health. *arXiv* **2022**, arXiv:2204.04777.
6. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 689–696.
7. Giuseppe, B. *Machine Learning Algorithms*; Packt Publishing Ltd.: Birmingham, UK, 2017.
8. Yann, L.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
9. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors* **2023**, *23*, 2112. [CrossRef]
10. Tadas, B.; Ahuja, C.; Morency, L. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 423–443.
11. Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* 2021, *in press*. [CrossRef]
12. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **2021**, *13*, 218. [CrossRef]
13. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. *Radiographics* **2017**, *37*, 505. [CrossRef]
14. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 1–21. [CrossRef] [PubMed]
15. Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. *Glob. Transitions Proc.* **2021**, *2*, 24–28. [CrossRef]
16. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94. [CrossRef]
17. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381. [CrossRef]
18. Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Tirunelveli, India, 4–6 February 2021; pp. 1529–1534.
19. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
20. Larrañaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. *Industrial Applications of Machine Learning*; CRC Press: Boca Raton, FL, USA, 2018.
21. L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. *IEEE Access* **2017**, *5*, 7776–7797. [CrossRef]
22. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [CrossRef]
23. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems: Applications, challenges, and opportunities. *Artif. Intell. Rev.* **2021**, *54*, 3299–3348. [CrossRef]
24. Leskovec, J.; Rajaraman, A.; Ullman, J.D. *Mining of Massive Data Sets*; Cambridge University Press: Cambridge, UK, 2020.
25. Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in deploying machine learning: A survey of case studies. *ACM Comput. Surv.* **2020**, *55*, 1–29. [CrossRef]
26. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981. [CrossRef]
27. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]
28. Rosario, M.; Mukhopadhyay, S.C.; Liu, Z.; Slomovitz, D.; Samantaray, S.R. Advances on sensing technologies for smart cities and power grids: A review. *IEEE Sens. J.* **2017**, *17*, 7596–7610.

29. Total Data Volume Worldwide 2010–2025 | Statista. Petroc Taylor. 8 September 2022. Statista. Available online: https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed on 15 February 2023).
30. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [CrossRef]
31. Lidong, W. Heterogeneous data and big data analytics. *Autom. Control. Inf. Sci.* **2017**, *3*, 8–15.
32. Geert, L.; Ciompi, F.; Wolterink, J.M.; de Vos, B.D.; Leiner, T.; Teuwen, J.; Išgum, I. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc. Imaging* **2019**, *12*, 1549–1565.
33. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review. *Sensors* **2023**, *23*, 828. [CrossRef]
34. Aamir, J.; Zghyer, F.; Kim, C.; Spaulding, E.M.; Isakadze, N.; Ding, J.; Kargillis, D.; Gao, Y.; Rahman, F.; Brown, D.E.; et al. Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology. *Am. J. Prev. Cardiol.* **2022**, *12*, 100379.
35. Ramesh, A.N.; Kambhampati, C.;Monson, J.R.L.; Drew, P.J. Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* **2004**, *86*, 334. [CrossRef]
36. Maddox, T.M.; Rumsfeld, J.S.; Payne, P.R. Questions for artificial intelligence in health care. *JAMA* **2019**, *321*, 31–32. [CrossRef]
37. Amine, M.M.; Adda, M.; Bouzouane, A.; Ibrahim, H. Machine learning and smart devices for diabetes management: Systematic review. *Sensors* **2022**, *22*, 1843. [CrossRef]
38. Shweta, C.; Biswas, N.; Jones, L.D.; Kesari, S.; Ashili, S. Smart Consumer Wearables as Digital Diagnostic Tools: A Review. *Diagnostics* **2022**, *12*, 2110. [CrossRef]
39. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472. [CrossRef]
40. Yukang, X. A review on intelligent wearables: Uses and risks. *Hum. Behav. Emerg. Technol.* **2019**, *1*, 287–294.
41. Marie, C.; Estève, D.; Fourniols, J.; Escriba, C.; Campo, E. Smart wearable systems: Current status and future challenges. *Artif. Intell. Med.* **2012**, *56*, 137–156.
42. Chinthaka, J.S.M.D.A.; Ganegoda, G.U. Involvement of machine learning tools in healthcare decision making. *J. Healthc. Eng.* **2021**, *2021*, 6679512.
43. Sameer, Q. Artificial intelligence and machine learning in precision and genomic medicine. *Med. Oncol.* **2022**, *39*, 120.
44. Arjun, P. *Machine Learning and AI for Healthcare*; Apress: Coventry, UK, 2019.
45. Nitish, S.; Salakhutdinov, R.R. Multimodal learning with deep boltzmann machines. *Adv. Neural Inf. Process. Syst.* **2014**, *15*, 2949–2980.
46. White, F.E. *Data Fusion Lexicon*; Joint Directors of Labs: Washington, DC, USA, 1991.
47. Baronio, M.A.; Cazella, S.C. Multimodal Deep Learning for Computer-Aided Detection and Diagnosis of Cancer: Theory and Applications. *Enhanc. Telemed. Health Adv. Iot Enabled Soft Comput. Framew.* **2021**, 267–287.
48. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Challenges and applications in multimodal machine learning. In *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition—Volume 2*; Association for Computing Machinery and Morgan & Claypool: San Rafael, CA, USA, 2018; pp. 17–48. [CrossRef]
49. Anil, R.; Walambe, R.; Ramanna, S.; Kotecha, K. Multimodal co-learning: Challenges, applications with datasets, recent advances and future directions. *Inf. Fusion* **2022**, *81*, 203–239.
50. Grigorios, T.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehous. Min.* **2007**, *3*, 1–13.
51. Min-Ling, Z.; Zhou, Z. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1819–1837.
52. Xibin, D.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258.
53. Omer, S.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249.
54. Announcing MMF: A Framework for Multimodal AI Models. Available online: https://ai.facebook.com/blog/announcing-mmf-a-framework-for-multimodal-ai-models/ (accessed on 18 February 2023).
55. Hasib-Al, R.; Ovi, P.R.; Gangopadhyay, A.; Mohsenin, T. TinyM2Net: A Flexible System Algorithm Co-designed Multimodal Learning Framework for Tiny Devices. *arXiv* **2022**, arXiv:2202.04303.
56. Pengcheng, X.; Shu, C.; Goubran, R. A Unified Deep Learning Framework for Multi-Modal Multi-Dimensional Data. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019; pp. 1–6.
57. Ma, S.L.R.Y.; Zeng, C.; Boussioux, L.; Carballo, K.V.; Na, L.; Wiberg, H.M.; Li, M.L.; Fuentes, I.; Bertsimas, D. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit. Med.* **2022**, *5*, 149.
58. Hang, L.; Kang, Y.; Hao, Y.; Ding, W.; Wu, Z.; Liu, Z. A Multimodal Machine Learning Framework for Teacher Vocal Delivery Evaluation. In *Proceedings of the Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, 14–18 June 2021*; Springer International Publishing: Cham, Switzerland, 2021; Part II, pp. 251–255.
59. Valerio, B.; Ceravolo, P.; Maghool, S.; Siccardi, S. Toward a general framework for multimodal big data analysis. *Big Data* **2022**, *10*, 408–424.
60. YJing, A.; Liang, N.; Pitts, B.J.; Prakah-Asante, K.O.; Curry, R.; Blommer, M.; Swaminathan, R.; Yu, D. Multimodal Sensing and Computational Intelligence for Situation Awareness Classification in Autonomous Driving. *IEEE Trans. Hum.-Mach. Syst.* **2023**, *53*, 270–281.

61. Azin, A.; Saha, R.; Jakubovitz, D.; Peyre, J. AutoFraudNet: A Multimodal Network to Detect Fraud in the Auto Insurance Industry. *arXiv* **2023**, arXiv:2301.07526.
62. Arnab, B.; Ahmed, M.U.; Begum, S. A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. *IEEE Access* **2023**, 11, 14804–14831.
63. Lemay, P.C.S.D.G.; Owen, C.L.; Woodward-Greene, M.J.; Sun, J. Multimodal AI to Improve Agriculture. *IT Prof.* **2021**, *23*, 53–57.
64. Yuchen, Z.; Barnaghi, P.; Haddadi, H. Multimodal federated learning on iot data. In Proceedings of the 2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI), Milano, Italy, 4–6 May 2022; pp. 43–54.
65. Min, C.; Hao, Y.; Hwang, K.; Wang, L.; Wang, L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access* **2017**, *5*, 8869–8879.
66. Bersche, G.S.; Shibahara, T.; Agboola, S.; Otaki, H.; Sato, J.; Nakae, T.; Hisamitsu, T.; Kojima, G.; Felsted, J.; Kakarmath, S.; et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. *Bmc Med. Inform. Decis. Mak.* **2018**, *18*, 1–17.
67. Yuan, L.; Mao, C.; Yang, Y.; Wang, F.; Ahmad, F.S.; Arnett, D.; Irvin, M.R.; Shah, S.J. Integrating hypertension phenotype and genotype with hybrid non-negative matrix factorization. *Bioinformatics* **2019**, *35*, 1395–1403.
68. Rao, W.R.R.D.C.; Ellison, R.C.; Arnett, D.K.; Heiss, G.; Oberman, A.; Eckfeldt, J.H.; Leppert, M.F.; Province, M.A.; Mockrin, S.C.; Hunt, S.C.; et al. NHLBI family blood pressure program: Methodology and recruitment in the HyperGEN network. *Ann. Epidemiol.* **2000**, *10*, 389–400.
69. Xiaosong, W.; Peng, Y.; Lu, L.; Lu, Z.; Summers, R.M. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9049–9058.
70. Xiaosong, W.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
71. Dina, D.; Kohli, M.D.; Rosenman, M.B.; Shooshan, S.E.; Rodriguez, L.; Antani, S.; Thoma, G.R.; McDonald, C.J. Preparing a collection of radiology examinations for distribution and retrieval. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 304–310.
72. Yixue, H.; Usama, M.; Yang, J.; Hossain, M.S.; Ghoneim, A. Recurrent convolutional neural network based multimodal disease risk prediction. *Future Gener. Comput. Syst.* **2019**, *92*, 76–83.
73. Yikuan, L.; Wang, H.; Luo, Y. A comparison of pre-trained vision-and-language models for multimodal representation learning across medical images and reports. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 1999–2004.
74. Pollard, J.A.E.W.T.J.; Greenbaum, N.R.; Lungren, M.P.; Deng, C.; Peng, Y.; Lu, Z.; Mark, R.G.; Berkowitz, S.J.; Horng, S. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv* **2019**, arXiv:1901.07042.
75. Ilan, S.; Cirulli, E.T.; Huang, L.; Napier, L.A.; Heister, R.R.; Hicks, M.; Cohen, I.V.; Yu, H.C.; Swisher, C.L.; Schenker-Ahmed, N.M.; et al. An unsupervised learning approach to identify novel signatures of health and disease from multimodal data. *Genome Med.* **2020**, *12*, 1–14.
76. Esra, Z.; Madai, V.I.; Khalil, A.A.; Galinovic, I.; Fiebach, J.B.; Kelleher, J.D.; Frey, D.; Livne, M. Multimodal Fusion Strategies for Outcome Prediction in Stroke. In Proceedings of the 13th International Conference on Health Informatics, Valletta, Malta, 24–26 February 2020; pp. 421–428.
77. Benjamin, H.; Pittl, S.; Ebinger, M.; Oepen, G.; Jegzentis, K.; Kudo, K.; Rozanski, M.; Schmidt, W.U.; Brunecker, P.; Xu, C.; et al. Prospective study on the mismatch concept in acute stroke patients within the first 24 h after symptom onset-1000Plus study. *BMC Neurol.* **2009**, *9*, 60.
78. Shih-Cheng, H.; Pareek, A.; Zamanian, R.; Banerjee, I.; Lungren, M.P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: A case-study in pulmonary embolism detection. *Sci. Rep.* **2020**, *10*, 1–9.
79. Ayoub, B.; Groenhof, T.K.J.; Veldhuis, W.B.; de Jong, P.A.; Asselbergs, F.W.; Oberski, D.L. Multimodal learning for cardiovascular risk prediction using EHR data. *arXiv* **2020**, arXiv:2008.11979.
80. Gerarda, S.P.C.; Algra, A.; Laak, M.F.V.D.; Grobbee, D.E.; Graaf, Y.V.D. Second manifestations of ARTerial disease (SMART) study: rationale and design. *Eur. J. Epidemiol.* **1999**, *15*, 773–781.
81. Gianluca, B.; Neuberger, U.; Mahmutoglu, M.A.; Foltyn, M.; Herweh, C.; Nagel, S.; Schönenberger, S.; Heiland, S.; Ulfert, C.; Ringleb, P.A.; et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke* **2020**, *51*, 3541–3551.
82. Makoto, N.; Kiuchi, K.; Nishimura, K.; Kusano, K.; Yoshida, A.; Adachi, K.; Hirayama, Y.; Miyazaki, Y.; Fujiwara, R.; Sommer, P.; et al. Accessory pathway analysis using a multimodal deep learning model. *Sci. Rep.* **2021**, *11*, 8045.
83. Larry, H.; Kim, R.; Tokcan, N.; Derksen, H.; Biesterveld, B.E.; Croteau, A.; Williams, A.M.; Mathis, M.; Najarian, K.; Gryak, J. Multimodal tensor-based method for integrative and continuous patient monitoring during postoperative cardiac care. *Artif. Intell. Med.* **2021**, *113*, 102032.
84. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability. *Procedia Comput. Sci.* **2022**, *203*, 231–238.
85. Matthew, T. Gesture recognition. In *Handbook of Virtual Environments*; CRC Press: Boca Raton, FL, USA, 2002; pp. 263–278.

86. Armando, P.; Mital, M.; Pisano, P.; Giudice, M.D. E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation. *Technol. Forecast. Soc. Chang.* **2020**, *153*, 119226.
87. Nasiri, A.Z.; Rahmani, A.M.; Hosseinzadeh, M. The role of the Internet of Things in healthcare: Future trends and challenges. *Comput. Methods Programs Biomed.* **2021**, *199*, 105903.
88. K, D.S.; Kory, J. A review and meta-analysis of multimodal affect detection systems. *Acm Comput. Surv.* **2015**, *47*, 1–36.
89. Yoshua, B.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828.
90. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. L. Rev.* **2016**, *2*, 287. [CrossRef]
91. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98. [CrossRef]
92. Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. *Am. J. Comp. Law* **1986**, *34*, 715–743. [CrossRef]
93. Chik, W.B. The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. *Comput. Law Secur. Rev.* **2013**, *29*, 554–575. [CrossRef]
94. Islam, M.K.; Rastegarnia, A.; Sanei, S. Signal Artifacts and Techniques for Artifacts and Noise Removal. In Signal Processing Techniques for Computational Health Informatics; Springer: Cham, Switzerland, 2021; pp. 23–79.
95. Daly, I.; Billinger, M.; Scherer, R.; Müller-Putz, G. On the automated removal of artifacts related to head movement from the EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 427–434. [CrossRef]
96. Dyk, V.; A, D.; Meng, X. The art of data augmentation. *J. Comput. Graph. Stat.* **2001**, *10*, 1–50.
97. Dalwinder, S.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524.
98. Jameel, M.A.; Hassan, M.M.; Kadir, D.H. Improving classification performance for a novel imbalanced medical dataset using SMOTE method. *Int. J.* **2020**, *9*, 3161–3172.
99. Ganesh, K.; Basri, S.; Imam, A.A.; Khowaja, S.A.; Capretz, L.F.; Balogun, A.O. Data harmonization for heterogeneous datasets: A systematic literature review. *Appl. Sci.* **2021**, *11*, 8275.
100. Michal, S.G.; Rubinfeld, D.L. Data standardization. *NYUL Rev.* **2019**, *94*, 737.
101. Maksymilian, W.; Chen, K. Feature importance ranking for deep learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 5105–5114.
102. Angelos, C.; Martins, R.M.; Jusufi, I.; Kerren, A. A survey of surveys on the use of visualization for interpreting machine learning models. *Inf. Vis.* **2020**, *19*, 207–233.
103. Alberto, B.; Domingo-Ferrer, J. Machine learning explainability through comprehensible decision trees. In *Machine Learning and Knowledge Extraction: Third IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2019, Canterbury, UK, 26–29 August 2019*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 15–26.
104. Stephan, W. *Towards Explainable Artificial Intelligence: Interpreting Neural Network Classifiers with Probabilistic Prime Implicants*; Technische Universitaet: Berlin, Germany, 2022.
105. Edwin, L. Evolving fuzzy and neuro-fuzzy systems: Fundamentals, stability, explainability, useability, and applications. In *Handbook on Computer Learning and Intelligence: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*; World Scientific: Singapore, 2022; pp. 133–234.
106. Shima, K.; Eftekhari, M. Feature selection using multimodal optimization techniques. *Neurocomputing* **2016**, *171*, 586–597.
107. Tejalal, C.; Mishra, V.; Goswami, A.; Sarangapani, J. A comprehensive survey on model compression and acceleration. *Artif. Intell. Rev.* **2020**, *53*, 5113–5155.
108. Shiliang, S.; Cao, Z.; Zhu, H.; Zhao, J. A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern. 2019*, **50**, 3668–3681.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.