




Unsupervised Learning in Precision Medicine: Unlocking Personalized Healthcare through AI

Alfonso Trezza ^{1,†}, Anna Visibelli ^{1,†} , Bianca Roncaglia ^{1,†} , Ottavia Spiga ^{1,2,3}  and Annalisa Santucci ^{1,3,*} 

¹ Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy; alfonso.trezza2@unisi.it (A.T.); anna.visibelli2@unisi.it (A.V.); bianca.roncaglia@unisi.it (B.R.)

² Centro della Scienza e della Tecnica, Polo Universitario Grossetano, Via Ginori 41, 58100 Grosseto, Italy

³ Competence Center ARTES 4.0, 53100 Siena, Italy

* Correspondence: annalisa.santucci@unisi.it

† These authors contributed equally to this work.

Abstract: Integrating Artificial Intelligence (AI) into Precision Medicine (PM) is redefining healthcare, enabling personalized treatments tailored to individual patients based on their genetic code, environment, and lifestyle. AI's ability to analyze vast and complex datasets, including genomics and medical records, facilitates the identification of hidden patterns and correlations, which are critical for developing personalized treatment plans. Unsupervised Learning (UL) is particularly valuable in PM as it can analyze unstructured and unlabeled data to uncover novel disease subtypes, biomarkers, and patient stratifications. By revealing patterns that are not explicitly labeled, unsupervised algorithms enable the discovery of new insights into disease mechanisms and patient variability, advancing our understanding of individual responses to treatment. However, the integration of AI into PM presents some challenges, including concerns about data privacy and the rigorous validation of AI models in clinical practice. Despite these challenges, AI holds immense potential to revolutionize PM, offering a more personalized, efficient, and effective approach to healthcare. Collaboration among AI developers and clinicians is essential to fully realize this potential and ensure ethical and reliable implementation in medical practice. This review will explore the latest emerging UL technologies in the biomedical field with a particular focus on PM applications and their impact on human health and well-being.

Keywords: artificial intelligence; precision medicine; healthcare



Citation: Trezza, A.; Visibelli, A.; Roncaglia, B.; Spiga, O.; Santucci, A. Unsupervised Learning in Precision Medicine: Unlocking Personalized Healthcare through AI. *Appl. Sci.* **2024**, *14*, 9305. <https://doi.org/10.3390/app14209305>

Academic Editor: Lucia Billeci

Received: 24 September 2024

Revised: 3 October 2024

Accepted: 8 October 2024

Published: 12 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI) is transforming healthcare, particularly from a Precision Medicine (PM) perspective, the branch of medicine that focuses on tailoring medical treatments to individual patients based on unique genetic, environmental, and lifestyle factors. This personalized approach represents a major shift from the traditional “one-size-fits-all” paradigm: the ability of AI to process and analyze vast and complex datasets, such as genomic data, medical histories, and patient health information, is critical to creating targeted diagnostic and therapeutic methods. The “five Vs” of big data—volume, velocity, variety, veracity, and value—highlight the complexities that healthcare data present, requiring sophisticated AI models to manage these dimensions effectively. As reported in Johnson et al. [1], the convergence of AI and PM is already demonstrating significant potential in areas such as early disease detection, treatment planning, and genomics-guided therapies. AI's power to integrate structured and unstructured data, along with its capabilities in natural language processing, computer vision, and conversational systems, enables it to generate actionable insights to support clinical decision-making. The societal impact of AI in healthcare is predicted to expand dramatically. According to a recent Elsevier report [2], 72% of researchers and physicians think AI will revolutionize healthcare within the next few years. However, this transformative potential comes with challenges, including concerns

over data privacy, model reliability, and the risk of bias in AI outputs. As noted in both reports, data security and trust are critical for the successful integration of AI systems in healthcare. A key aspect of AI in this context is unsupervised learning (UL), which can identify hidden patterns within unlabeled data, generating new insights that are crucial for disease stratification, biomarker identification, and patient classification. But as these technologies become more integrated into healthcare, guaranteeing the fairness, transparency, and ethical application of AI systems continues to be a crucial priority. This review will explore the latest developments in UL, its applications in PM, and the future directions for integrating AI-driven solutions into personalized care. We will also address the ethical challenges associated with AI, the need for more robust data security frameworks, and the potential impact of these technologies on patient outcomes and healthcare costs.

2. Artificial Intelligence

The first modern definition of AI dates back to 1956 and is attributed to Professor John McCarthy (Stanford University), who described AI as the science and engineering of creating intelligent machines. Paraphrasing McCarthy, AI can be defined as “*the use of computers and technology to simulate intelligent behavior and critical thinking comparable to that of a human being*”. Early foundational contributions began as far back as the 18th century with pioneers like Basile Bouchon and Charles Babbage, whose mechanical inventions laid the groundwork for modern computing [3,4]. By the 1940s and 1950s, researchers such as Warren McCulloch, Walter Pitt, and others developed the first artificial neurons and neural networks, pushing the field forward. Early computers were particularly adept at solving mathematically driven tasks, which are intellectually challenging for humans but relatively simple for machines, as they can be described by a set of formal mathematical rules. However, the real challenge for AI emerged in replicating perceptual skills that humans have developed over hundreds of thousands of years through evolution. This challenge led to the advent of machine learning (ML), and more recently, deep learning [5], which mimics the layered learning process of the human brain. In a general frame, AI models are capable of acquiring, storing, and using data—known as training data—to learn and enhance their performance [6]. The way these data are gathered, processed, and labeled determines the key differences between various types of AI models, particularly those within ML. An AI algorithm typically receives data either provided by developers or collected autonomously, which it then uses to expand its knowledge base and perform tasks more effectively. There are three main categories of AI algorithms: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL) [7]. Each of these categories works based on distinct principles, which influence their applicability in various fields, including Precision Medicine (PM).

- **Supervised Learning:** In this approach, the training data are labeled with a target, representing the “expected result”. After the training phase, the system can use the learned information to address problems that involve similar foundational knowledge.
- **Unsupervised Learning:** This method operates on an unlabeled training set, focusing on discovering patterns and relationships within the data without any prior knowledge about its structure or categories.
- **Reinforcement Learning:** Unlike the other methods, reinforcement learning also uses an unlabeled training set but provides feedback in the form of positive or negative results. This feedback creates a loop that allows the algorithm to assess whether its proposed solutions effectively resolve a problem, resembling the human learning process through “trial and error”.

The primary differences between these algorithms lie in how they are trained to learn from data, and consequently, in how they operate. While SL relies on labeled datasets to make predictions, UL focuses on uncovering hidden structures without prior labels, enabling it to handle complex, high-dimensional data typical in PM. RL, on the other hand, emphasizes learning through interaction with the environment, refining its strategies based on the outcomes of its actions. While SL and RL have their strengths, UL’s ability to analyze

unstructured data without prior labeling makes it a powerful tool in PM. For this reason, this review will introduce the most popular and common UL models and explain where they are typically employed in the field of PM.

2.1. AI Algorithms: Introduction to Unsupervised Learning

UL is a crucial branch of ML that focuses on algorithms using data that lack explicit labels or predefined outcomes [8]. This approach contrasts with SL, where models are trained on labeled data with known input–output pairs [9]. In UL, the system is designed to autonomously explore and analyze the structure of the data, allowing it to uncover hidden patterns, relationships, and features that may not be immediately recognizable [10] (see Figure 1).

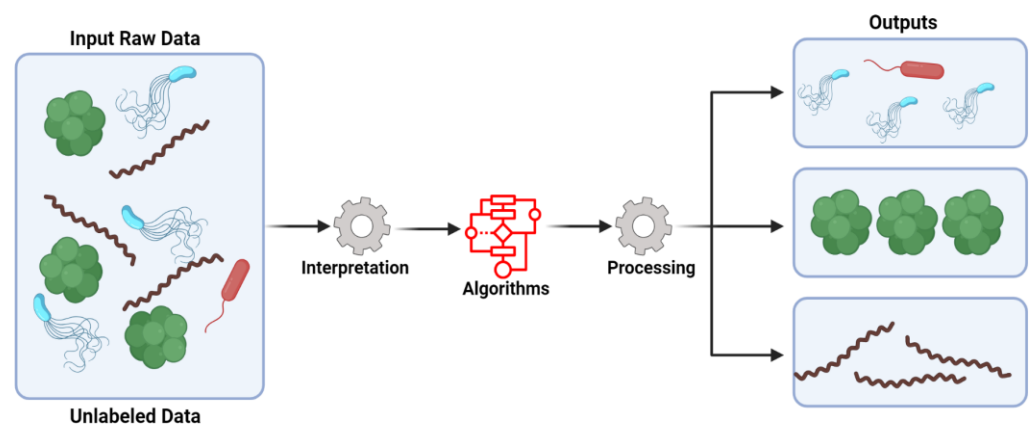


Figure 1. The workflow of the unsupervised learning model.

These hidden structures may reveal novel insights into the relationships between data points, enabling researchers to make discoveries that conventional analysis was unable to reveal. This type of learning is particularly well-suited for complex data processing tasks where manual labeling of data is impossible due to the complexity of the dataset. For example, in healthcare, UL can be used to uncover new disease subtypes or predict patient outcomes based on unstructured electronic health records and genetic data [11]. One of the primary applications of UL is clustering [12], where the algorithm splits large datasets into subgroups based on inherent similarities. Clustering methods are commonly employed in a variety of fields, including genomics [13] and image analysis, to segment data into coherent categories without prior knowledge of their structure [14]. In addition to clustering, UL is also effective in dimensionality reduction, where algorithms are used to simplify high-dimensional data while preserving their most important features [15]. This process is crucial in fields such as bioinformatics, where datasets often contain thousands of variables, making it challenging to extract meaningful insights without reducing their complexity. This learning category also plays a central role in anomaly detection, a task that involves identifying data points that deviate significantly from the norm [16]. Outliers are critical in medical imaging to detect anomalies, such as cancers, that radiologists might not be able to recognize [17]. Despite its strengths, since there are no labeled outcomes to guide the learning process, evaluating the performance of unsupervised models can be challenging. In many cases, domain-specific knowledge is required to interpret the results and determine whether the identified patterns are meaningful [18]. An in-depth analysis of existing unsupervised models, discussing their methodologies, applications, and potential limitations within various research contexts, is therefore necessary.

2.2. Unsupervised Learning: Clustering

Cluster analysis can be categorized into three main objectives. Among the various clustering methods, K-means clustering [19] and hierarchical clustering [20] are two of the most widely implemented approaches, as shown in Figure 2.

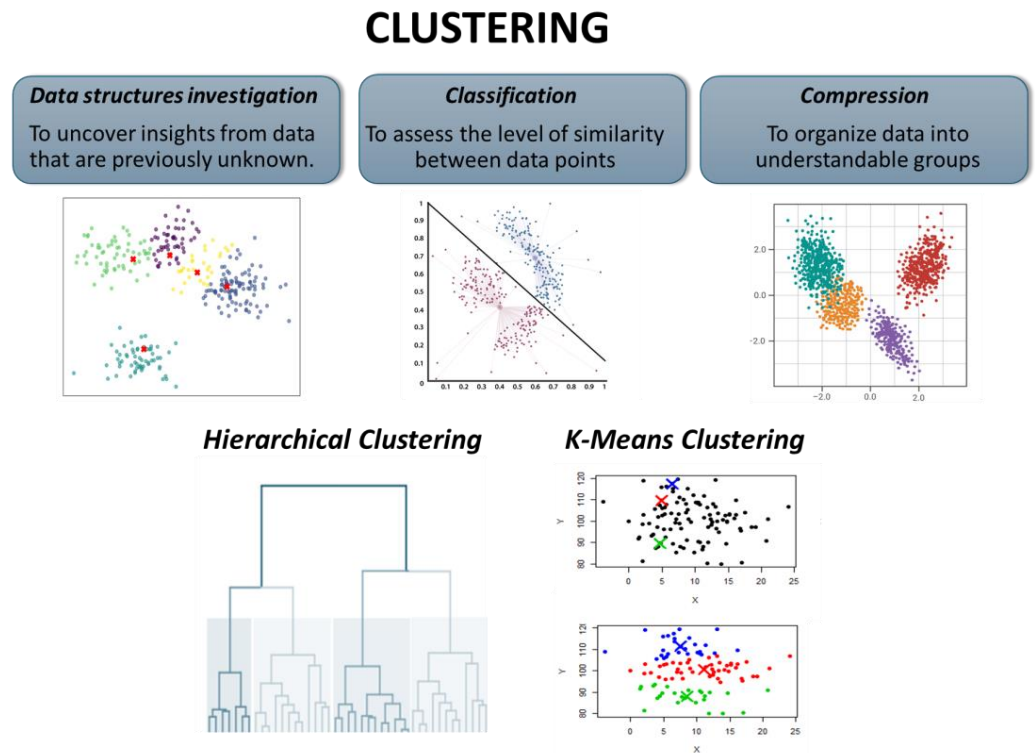


Figure 2. Key objectives and visualization models of cluster analysis are presented. At the top, data structures investigation, classification, and compression are illustrated. Below, hierarchical and K-means clustering techniques are displayed.

2.2.1. K-Means Clustering

K-means clustering [19] is a widely used partition-based clustering technique designed to segment a dataset into k distinct clusters. K-means aims to minimize the sum of squared distances between each data point and its assigned cluster centroid, thus the geometric centers of data points in a given space. Let $X = \{x_1, x_2, \dots, x_n\}$ represent the dataset with n data points. The goal is to partition the dataset into k clusters C_1, C_2, \dots, C_k and minimize the following objective function:

$$\sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - c_i\|^2$$

where c_i is the centroid of cluster C_i , and $\|\cdot\|$ represents the Euclidean distance.

The process begins with the algorithm randomly selecting k initial cluster centroids. Subsequently, each data point is assigned to the nearest centroid, resulting in the formation of k clusters based on a distance metric. The centroids of these clusters are then recalculated as the mean of all data points assigned to each cluster. This assignment and update process is iteratively repeated until the centroids stabilize, indicating that the algorithm has converged. The K-means algorithm is particularly suitable for handling large datasets with many features [21]. However, K-means needs the number of clusters specified in advance, which can be challenging if the optimal number of clusters is unknown.

2.2.2. Hierarchical Clustering

Hierarchical clustering [20] builds a hierarchy of clusters through two primary approaches: agglomerative (bottom-up) and divisive (top-down). In the agglomerative approach, the process begins with each data point as its cluster. These clusters are then iteratively merged based on a chosen distance metric. The process must be repeated until the desired number of clusters is achieved. This method can be described mathematically as follows:

- **Step 1.** Calculate pairwise distances between all clusters: Let $X = \{x_1, x_2, \dots, x_n\}$ represent the dataset with n data points, where each x_i is in its cluster C_i . The distance between two clusters $d(C_i, C_j)$ can be defined using different linkage methods.
- **Step 2.** Merge the Closest Clusters: find the two clusters C_i, C_j with the smallest distance $d(C_i, C_j)$ and merge them into a new cluster: $C_{ij} = C_i \cup C_j$.
- **Step 3.** Update the Distance Matrix: after merging, update the distance matrix to reflect the new distances between the merged cluster C_{ij} and all remaining clusters.

The divisive approach, in contrast, starts with all data points grouped into a single cluster. This cluster is then split into progressively smaller clusters until the desired number of clusters is obtained. Compared to K-means, hierarchical clustering does not require specifying the number of clusters in advance, which provides flexibility in determining the appropriate number of clusters. However, the computational complexity of hierarchical clustering can be substantial, especially with large datasets, making it less scalable than K-means.

2.3. Unsupervised Learning: Dimensionality Reduction

Dimensionality reduction is a crucial process in data analysis and ML, aimed at reducing the number of features in a dataset while considering only the relevant information. This process is essential for simplifying models, enhancing computational efficiency, and addressing challenges related to high-dimensional data. Among the various techniques available, see Figure 3, Principal Component Analysis (PCA) [22], t-Distributed Stochastic Neighbor Embedding (t-SNE) [23], and autoencoders [24] are prominent methods used for dimensionality reduction.

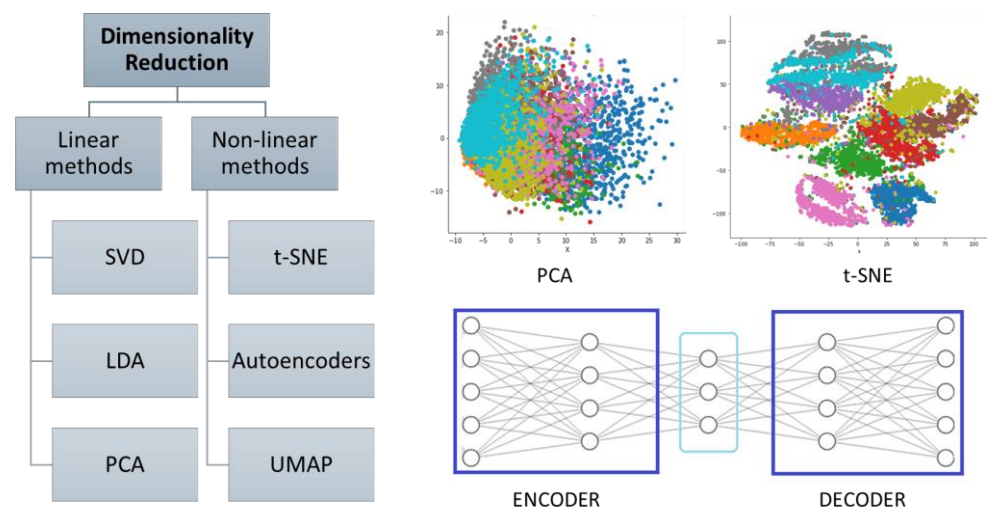


Figure 3. Linear and non-linear techniques for dimensionality reduction, including PCA, t-SNE, and autoencoders graphically represented on the right.

2.3.1. Principal Component Analysis

PCA [22] is a widely employed technique for linear dimensionality reduction that transforms the original data into a new coordinate system. PCA is effective in reducing dimensionality while preserving the most significant variance in the dataset. In this new system, the axes, known as principal components, are orthogonal directions that capture the maximum variance present in the data. PCA first computes the eigenvectors and eigenvalues of the data's covariance matrix. The eigenvectors determine the directions of maximum variance, while the eigenvalues represent the magnitude of this variance. By projecting the original data onto these principal components, PCA creates a reduced-dimensional representation that maintains most of the data's variability. The process can be summarized as follows:

- **Step 1:** center the dataset by subtracting the mean from each feature.
- **Step 2:** calculate the covariance matrix Σ of the centered data X_c :

$$\Sigma = \frac{1}{n-1} X_c^T X_c$$

- **Step 3:** compute the eigenvalues λ and eigenvectors v the covariance matrix:

$$\Sigma v = \lambda v$$

- **Step 4:** select the top k eigenvectors corresponding to the largest eigenvalues and project the data onto these eigenvectors to obtain the reduced-dimensional representation X_r :

$$X_r = X_c V_k$$

where V_k is the matrix containing the top k eigenvectors.

Moreover, this method simplifies the data, improves computational efficiency, and tackles multicollinearity. However, it assumes linear relationships, which may not capture non-linear structures, limiting its effectiveness in such cases.

2.3.2. t-Distributed Stochastic Neighbor Embedding

t-SNE [23] is a non-linear dimensionality reduction method specifically designed to visualize high-dimensional data in lower dimensions. The method focuses on preserving the local data structure, ensuring that similar data points are positioned close to each other in the reduced-dimensional space. Using probabilities, t-SNE calculates pairwise similarities between data points in the high-dimensional space. The approach seeks to minimize the divergence between the probability distributions of the high-dimensional and low-dimensional spaces, modeling these similarities as conditional probabilities. The optimization process in t-SNE involves gradient descent to adjust the placement of data points in the lower-dimensional space. This iterative refinement aims to maintain the local neighborhood relationships, allowing for a meaningful representation of the data structure. However, the algorithm can be computationally expensive, especially for large datasets, and its results are sensitive to hyperparameters, requiring careful tuning.

2.3.3. Autoencoders

Autoencoders [24] are UL models used for dimensionality reduction and anomaly detection. These methods can capture complex, non-linear relationships in the data, which makes them particularly effective for handling high-dimensional and intricate datasets. The networks have two primary components: the encoder and the decoder. The encoder compresses the input data into a lower-dimensional representation, known as the latent space. This compression step captures the essential features of the data while significantly reducing their dimensionality. Then, the decoder reconstructs the original data from the lower-dimensional representation. The objective of the decoder is to minimize the reconstruction error, ensuring that the compressed data retain as much of the original information as possible. However, training these networks can require significant resources, especially when dealing with deep networks.

2.4. Unsupervised Learning: Anomaly Detection

Another key aspect of UL is anomaly detection, which identifies outliers in data without predefined labels. Two prominent techniques for this are the One-Class Support Vector Machine (SVM) [25] and autoencoders [26]. Autoencoders can detect anomalies by learning to reconstruct input data, where outliers are detected based on poor reconstruction.

One-Class Support Vector Machine

The One-Class Support Vector Machine (OCSVM) [25] is a variant of the SVM algorithm designed for the task of anomaly detection. The OCSVM is particularly effective in detecting outliers in high-dimensional spaces and can manage non-linear data distributions. It focuses on identifying the regions in the feature space where the majority of data points are concentrated, classifying any points that fall outside these regions as anomalies, as shown in Figure 4.

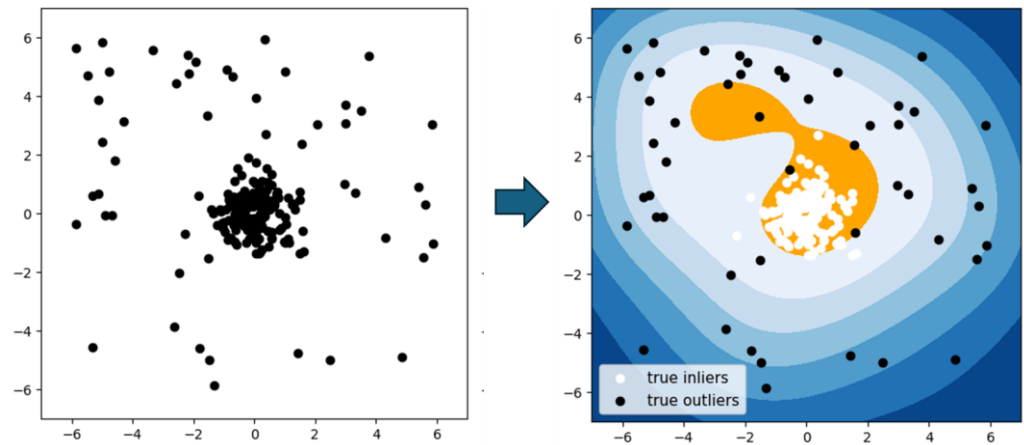


Figure 4. An example of OCSVM implementation for anomaly detection. The model learns the boundary of normal data and flags points outside this boundary as anomalies.

During the training phase, the OCSVM constructs a decision boundary that encompasses most of the data points and can effectively distinguish normal data from potential outliers. Once the training is complete, the model classifies new data points based on their position relative to the established boundary. Data points that lie outside this boundary are flagged as anomalies.

The choice between autoencoders and OCSVM for anomaly detection depends on the specific characteristics of the dataset and the nature of the anomalies. Autoencoders are preferable for high-dimensional data, where they can model non-linear relationships to capture anomalies. Because the model depends on data reconstruction, it may sometimes lead to false positives if anomalies closely resemble normal data. OCSVM, on the other hand, is often preferred when working with smaller datasets, particularly when the anomalies are distinct and the data structure is less complex. However, OCSVM is less effective in capturing non-linear relationships, which limits its applicability in more complex contexts.

3. Precision Medicine: Origin and History

Over the last two decades, PM has gained significant attention, focusing on selecting the most appropriate drug for a patient based on personal characteristics and individual differences [27]. Central to this approach is pharmacogenomics, the study of how genetic mapping relates to diseases, which helps identify the most effective drugs with minimal side effects [28]. Given that most diseases result from a complex interplay of genetic and environmental factors, PM offers clear advantages in tailoring treatment to individual patients affected by pathologies of different natures, like rare diseases [29–31]. While PM seems to be a modern development driven by technological advances, its origins date back over a thousand years to Avicenna (980–1037 AD), a prominent figure in traditional Persian medicine. In his seminal work, *The Canon of Medicine*, Avicenna introduced the core idea of PM, observing that “every drug will have different effects on different bodies and organs of a person” [32]. He also foreshadowed concepts like pharmacokinetics and pharmacodynamics, noting the variation in drug action over time across individuals. Avicenna recognized that a person’s temperament, a combination of factors such as metabolism, behavior, and mental state, was crucial in determining drug efficacy. He also considered sex, age, habits,

climate, occupation, and body structure, presaging what modern medicine now refers to as phenotype [33]. In addition, Avicenna anticipated the concept of genotype through his notion of innate temperament, which is influenced by environmental conditions and lifestyle, a precursor to modern epigenetics. In his detailed studies, Avicenna categorized temperament using ten main factors: facial and body color, skeletal structure, hair characteristics, sleep–wake pattern, bodily excretions (feces, urine, and sweat), behavior, mental states, and mood. These characteristics were then grouped into four basic qualities: coldness or warmth, dryness or moisture. These concepts led Avicenna to conclude that innate and acquired temperaments were important in targeted therapy (PM) because they influenced the effectiveness of treatment [33]. In his treatise, Avicenna reported several examples of different responses to drugs, such as scammonia resin (*Convolvulus scammonia* L.), which showed greater efficacy in hot-temperament patients than in cold-temperament patients. Avicenna also explored how season, climate, diet, and social factors affected drug efficacy. His work on drug toxicity and resistance led to recommendations for adjusting dosages, using alternative drugs, or combining therapies to enhance treatment outcomes. That is why Avicenna's groundbreaking insights into the influence of individual characteristics, lifestyle, and environment on pharmacotherapy reflect a deep understanding of concepts that form the basis of modern PM.

3.1. Precision Medicine: The Modern Birth

PM is a term that refers to the personalization of treatment for a subpopulation of people who differ in their susceptibility to developing a particular disease or response to a specific drug [33,34]. Originally, PM was called “Personalized Medicine” but because this term could be misinterpreted as a treatment aimed only at a single individual, it was replaced globally by the term “Precision Medicine” [35]. This is also known as “stratified medicine”, “targeted therapy”, and “deep phenotyping” [36]. Historically, the concept of PM dates back to the early 20th century, with pioneers such as Sir William Osler, who stated, “It is much more important to know what sort of a patient has a disease than what sort of a disease a patient has”. However, significant progress in this field was achieved with the discovery of double-helix DNA in 1953, the development of Sanger sequencing in 1977, and the Human Genome Project, launched in 1990, which led to the nearly complete sequencing of the human genome in 2003 and further updates after that [37]. These findings reveal that complex diseases are influenced by interactions among multiple genes. Individuals with similar genetic alterations may develop different disorders, while the same disease can result from mutations in different genes [38]. Despite the growing interest of the international scientific community, PM gained public awareness thanks to U.S. President Barack Obama, who in 2015 announced the PM initiative, declaring, “Tonight, I am launching a new Precision Medicine initiative to bring us closer to curing diseases such as cancer and diabetes”. This increase in interest reflects the importance of PM not only in the academic and clinical context but also in the public sphere.

3.2. Traditional Medicine vs. Precision Medicine

Traditional medicine (TM) takes a generic approach, in which a specific drug treats all patients with a specific disease [39]. However, this approach has limitations, as only a portion of patients respond to treatment, while a significant proportion do not respond or experience side effects [40]. Causes of these interindividual differences may include genetic variations, age, gender, addictions, race, ethnicity, drug interactions, comorbidities, and environmental factors. These variables not only result in wasted drugs but also in increased healthcare costs and dissatisfaction on the part of patients and physicians. PM seeks to overcome these limitations by considering individual variability in genetic, socio-environmental, and lifestyle factors to propose targeted therapies [40]. The goal is an accurate assessment of molecular, environmental, and behavioral factors that influence health and disease, thus leading to more accurate diagnosis, rational disease prevention strategy, treatment selection, and development and optimization of new therapies as shown

in Figure 5 [41]. The main difference between TM and PM lies in access to Big Data, or large volumes of data about sick and healthy patients. With rapid advances in molecular biology and genetic testing becoming faster and cheaper, it is now possible for researchers to collect large volumes of data [42]. These data, combined with clinical, pharmacological, and socioeconomic information, can be analyzed using advanced algorithms, allowing them to identify patterns of therapeutic efficacy and apply targeted treatments only to susceptible populations. Tools employed by PM include science-OMICS [43,44], pharmacomics [45,46], molecular dynamics approaches [47,48], Big Data [49–51], and AI [52–54].

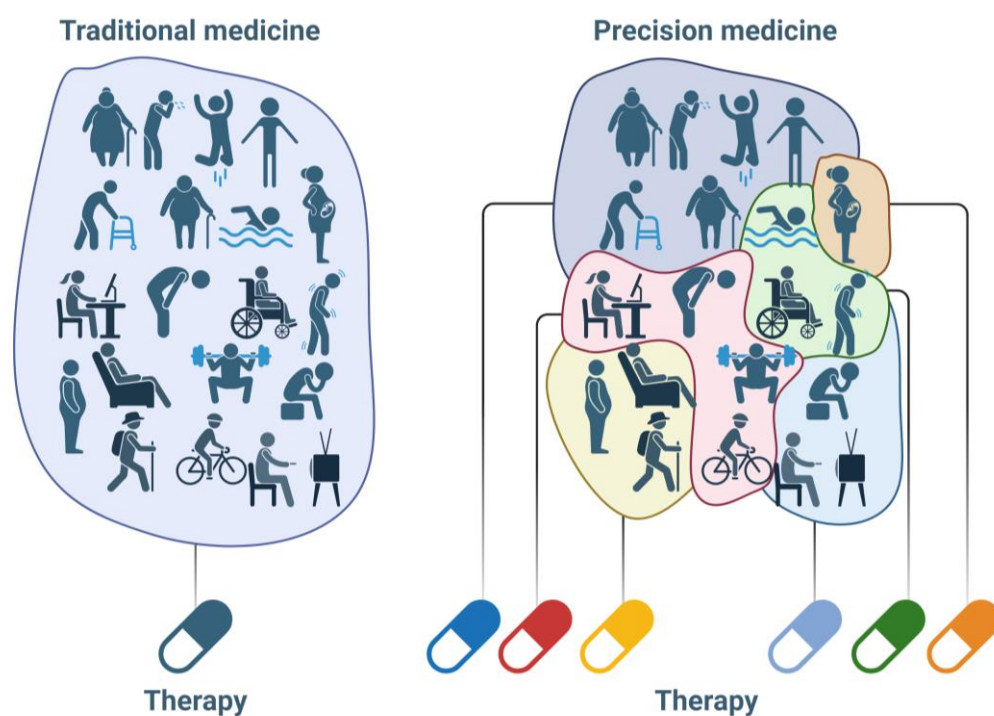


Figure 5. TM vs. PM. In the left image, a group of people, with different lifestyles but sharing the same pathology, can be observed. The approach of TM is to group the population under examination based on pathology and the provision of the same therapy for the entire population. In contrast, PM examines common lifestyles and genetic and/or environmental factors, identifying subgroups of populations, represented by the different circles, and confers a different therapy for each subgroup.

4. Ethical Considerations in AI-Driven PM

This data-driven approach raises critical concerns regarding patient privacy, especially when handling sensitive medical and genetic information. Given the high-stakes nature of PM, where personal health data are involved, robust solutions are required to ensure data protection without compromising the utility of AI in healthcare [55]. Privacy concerns are compounded by the vast amounts of data used by AI systems, such as electronic health records (EHRs), genomic data, and imaging. Ensuring compliance with privacy regulations, such as GDPR and HIPAA, while maintaining AI performance requires the adoption of advanced data security mechanisms such as encryption, anonymization, and access control. This limitation has driven a growing focus on privacy-preserving approaches, including Federated Learning (FL), blockchain technology, and generative adversarial networks [56].

FL, introduced by Google in 2016 [57], allows decentralized data analysis by enabling multiple institutions to collaboratively train ML models without centralizing the raw data. By sharing only model updates instead of patient data, FL enhances privacy and maintains the security of sensitive medical information. This approach ensures that ML can still leverage large-scale datasets to reveal crucial insights in PM while complying with privacy regulations like GDPR. To further strengthen privacy protections, additional mechanisms such as differential privacy and cryptographic methods (including homomorphic encryp-

tion and secure multi-party computation) are being integrated into newer FL models [58]. These techniques, when integrated with UL, form a robust framework for advancing PM. By protecting patient data while allowing for meaningful analysis, they address ethical concerns surrounding privacy and security. Moreover, cybersecurity measures, such as Secure Enclave technologies, should be implemented to safeguard data integrity and protect against cyber threats, which are increasingly pertinent as AI becomes more prevalent in healthcare. Another key consideration is transparency, as the complexity of UL models can obscure decision-making processes, raising skepticism among healthcare professionals and patients. Establishing guidelines for transparent reporting of UL models and their decision-making processes can build trust among clinicians and patients, ensuring that AI-generated insights are not only accurate but also interpretable and reliable. Bias in AI algorithms represents another significant ethical challenge [59]. If training datasets are not representative of diverse populations, the resulting model may inadvertently spread existing healthcare disparities. This could lead to unequal access to AI-driven treatments, and already marginalized groups may not benefit from the advancements in PM. So, mitigating bias in AI models necessitates diverse training datasets, fairness-aware development techniques [60], regular auditing, and ongoing evaluations to ensure equitable clinical applications. Finally, the integration of AI into PM, while offering promising prospects for personalizing care and accelerating research, raises important questions about professional responsibility and patient safety. Physicians must learn to view AI-generated recommendations as tools to support their clinical judgment and avoid delegating crucial decisions such as diagnosis or choice of therapy to these technologies [61]. The complexity of AI algorithms, which are often difficult to interpret, combined with concerns about data privacy and cybersecurity, makes it imperative to develop clear and binding guidelines to ensure ethical and responsible use of these technologies in healthcare. Investing in training healthcare professionals in the use of AI and promoting the development of more transparent and interpretable algorithms are key steps to fully exploit the potential of AI without compromising the quality and safety of care.

5. Unsupervised Learning Application in Precision Medicine

5.1. Clustering Application in PM

K-means clustering plays a crucial role in medical research by facilitating the identification of distinct patient subgroups, which can lead to more personalized treatment strategies and improved clinical outcomes. A recent study [62] employed K-means clustering to assess the immune signature of juvenile-onset systemic lupus erythematosus (SLE), aiming to enhance patient stratification. By selecting 8 subsets from 28 immune cell subsets, K-means categorizes patients into four distinct groups, revealing significant variations in T-cell frequencies. Correlation network analysis identified extensive immune correlations with clinical features, suggesting that K-means clustering can illuminate the multifactorial and heterogeneous nature of juvenile-onset SLE. The paper by Nelke et al. [63] describes the application of spherical K-means clustering for identifying distinct biological phenotypes in patients with acetylcholine receptor-antibody positive myasthenia gravis (MG). This UL method classified samples into different subgroups, using proteomics data. A modification of classical K-means was used, specifically spherical K-means, which employed cosine similarity as the distance metric. This was considered more suitable for high-dimensional data like proteomic profiles and offered better robustness against noise. Clustering revealed four distinct patient subgroups, with protein signature 3 (PS3) identified as having high disease severity, marked by complement activation, thus suggesting that PS3 patients might benefit from complement-inhibiting therapies. In the management of type 2 diabetes, K-means was used to classify patients based on genetic profiles, clinical biomarkers, and drug responses, allowing for more precise treatment strategies. One cluster, for example, showed higher insulin resistance, while another responded better to drugs like metformin, helping to reduce adverse reactions and long-term complications such as cardiovascular risks [64]. In acute ischemic stroke (AIS), over 1400 patients were K-means clustered

into groups based on metabolic and inflammatory markers, such as lipid levels and ion concentrations. One group, with high lipid levels, responded better to lipid-lowering drugs, while another cluster with elevated inflammation benefited from anti-inflammatory treatments. This approach personalized stroke care, improving recovery outcomes by predicting patient-specific trajectories [65]. In medical imaging, K-means was used to enhance MRI segmentation, grouping pixels by intensity and spatial data to detect subtle brain abnormalities, such as lesions in neurodegenerative diseases like multiple sclerosis. This method improved diagnostic accuracy, allowing for earlier intervention and more tailored treatment plans, significantly advancing patient outcomes compared to traditional techniques [66].

Hierarchical clustering offers flexibility in capturing data structure, making it ideal for exploratory analysis in complex, high-dimensional datasets. In a recent work of Neel S. Madhukar et al. [67], a hierarchical clustering algorithm was employed to develop the BANDIT tool, enabling it to uncover complex drug–target relationships and latent patterns within the global drug landscape that may be hidden in flat data representations. The efficacy of BANDIT was validated through extensive testing, where it successfully replicated known shared-target relationships, individual drug–target interactions, and established mechanisms of action within the test set. Furthermore, BANDIT’s predictions aligned with outcomes from large-scale experimental screens, demonstrating its robustness and reliability. Experimentally, several novel predictions made by BANDIT were confirmed using various bioassays and model systems, highlighting its potential to uncover previously unrecognized small molecules with therapeutic relevance, particularly in refractory tumors. The Bayesian framework further enhances BANDIT’s capabilities by allowing the seamless incorporation of new data, thereby continuously refining the predictive model. Analogously, the study of Henry Gerdes et al. applied systematically hierarchical clustering to analyze 466 drugs in acute myeloid leukemia (AML) and solid cancer cell lines, identifying numerous phosphorylation sites, proteins, and transcript markers associated with drug responses. Hierarchical clustering revealed that drugs with similar mechanisms are often grouped [68]. This robust application of hierarchical clustering not only facilitated the identification of drug response markers but also provided insights into the biological mechanisms driving these responses, highlighting its critical role in understanding drug actions in cancer therapy.

5.2. Dimensionality Reduction Application in PM

Dimensionality reduction plays a crucial role in PM. PCA has been employed in [69] to transform high-dimensional heart data into a lower-dimensional space, facilitating the study of genetic variations linked to anatomical traits like left ventricular morphology. However, the study also highlighted the limitations of PCA when applied to complex data, as it tends to overlook geometric details critical for accurate medical diagnostics. To address these limitations, the study compared PCA with the Convolutional Mesh Autoencoder (CoMA), which preserves topological and geometric complexity through spectral convolutions based on the Laplace–Beltrami operator. While PCA is limited to linear transformations, CoMA’s ability to handle non-linear variations enables more accurate reconstructions of cardiac shapes. This is particularly important for detecting subtle anatomical variations, which are often crucial for understanding the genetic basis of heart diseases. Another study [70] highlights PCA’s value in identifying neuroimaging patterns that predict cognitive decline, compressing high-dimensional relevance maps from brain scans of patients with mild cognitive impairment. By reducing the dimensionality of these maps, key components related to the progression of dementia were identified. The study retained 64 components that were significantly associated with cognitive decline. A Cox proportional hazard model demonstrated that these PCA-derived components could predict dementia prognosis over time, with several components correlating with cognitive impairments in domains such as language and executive function.

The t-SNE model has been particularly effective in visualizing and analyzing high-dimensional genomic and transcriptomic data, where traditional classification methods might struggle. In one study, t-SNE was applied to DNA methylation data to classify tumors [71]. By projecting cancer samples alongside a reference cohort, t-SNE helped to assign tumors to DNA methylation subclasses, and in cases of no classification, it revealed potential new subclasses. This approach, when integrated with additional molecular data, significantly enhanced tumor classification accuracy, contributing to more precise diagnostic and treatment plans. In another application [72], t-SNE was employed to analyze cellular diversity in multiple myeloma samples, clustering over 489,000 cells into 15 distinct groups using latent space data from a convolutional neural network. This dimensionality reduction allowed for the visualization of phenotypic landscapes, uncovering distinct bone marrow cell communities associated with different stages of the disease. The use of t-SNE, alongside proteotyping and DL, facilitated the identification of cell subpopulations, which aided in clinical decision-making and the assessment of disease progression. In colorectal cancer (CRC) research [73], t-SNE was used to downsample and visualize cells based on cluster size and marker expression, providing critical information about tumor-immune interactions and offering potential targets for therapeutic intervention. The method allowed for the differentiation of epithelial, stromal, and immune cells and identified biologically relevant subpopulations, such as budding tumor cells.

Although PCA has been widely used for dimensionality reduction in biomedical research, its linear structure makes it difficult to capture non-linear correlations that are present in biological datasets. Recent advances in deep learning have provided more powerful alternatives for dimensionality reduction, with techniques highly suitable for applications in PM, where high-dimensional datasets are becoming more and more prevalent. Autoencoders have emerged as powerful tools for non-linear dimensionality reduction, particularly in non-image medical data analysis. In one study, a hybrid autoencoder (HAE) framework was developed to improve dimensionality reduction for COVID-19 prognosis prediction [74]. Traditional methods like PCA and t-SNE, which are constrained by linearity or visual representation, often fall short of capturing complex relationships. In contrast, the HAE framework optimized the latent space by clustering similar data points more effectively, providing a lower-dimensional, informative representation of the raw medical data. This allowed for more efficient prognosis prediction, as the autoencoder uncovered non-linear patterns in the data that would otherwise remain hidden. Another application involved the use of deep autoencoders to analyze transcriptomic data for cancer diagnosis [75]. The DeepT2Vec autoencoder significantly reduced high-dimensional gene expression data into smaller feature vectors while retaining critical biological information. This compressed representation was employed in classification tasks, such as distinguishing between normal and tumor tissues. In the study of Wegmann et al. [76], deep learning-based dimensionality reduction, in the form of convolutional neural networks (CNNs), was employed to classify different cell types and states based on morphology and immunofluorescence data from malignant serous effusions (MSEs). These CNNs processed data from millions of individual cells, effectively identifying critical features related to cancer biology. By leveraging these techniques, the study could capture the diversity of cell types within MSE samples and correlate them with genomic and transcriptomic profiles. In another work [77], DL is applied to process high-dimensional mass spectrometry data, facilitating the identification of disease-specific metabolic profiles without the need for intermediate steps like peak extraction and annotation. The model uses a pre-pooling module to reduce the dimensionality of the raw MS data, transforming it from a three-dimensional structure into a two-dimensional matrix. The reduced data are passed through a CNN, which extracts features relevant to classifying disease states. This CNN-based feature extraction is critical for reducing the complexity of the data while maintaining its most important characteristics for identifying disease-related metabolic signals. By using an ensemble deep learning strategy, the method presented handles large and complex datasets across different hospitals, effectively reducing batch effects and enhancing classification accuracy.

5.3. Anomaly Detection Application in PM

In PM, anomaly detection algorithms can identify rare conditions or atypical responses to treatments, which is essential for developing personalized therapies. An excellent example is the use of autoencoders, commonly applied to detect anomalies in high-dimensional biological data. In a recent work [78], the context-aware deconfounding autoencoder (CODE-AE) was designed to address the out-of-distribution problem in PM. CODE-AE extracts meaningful biological signals and distinguishes confounding variables, effectively identifying anomalies in gene expression data. This enables more accurate prediction of patient-specific drug responses and improves the robustness of clinical outcome predictions from *in vitro* data. Another application focuses on developing a framework named Detect for anomaly detection in the brain microstructure using deep diffusion MRI (dMRI) tractometry [79]. The method uses autoencoders to learn normative patterns from healthy brain data and then detect deviations in individual patients without needing prior diagnostic labels. By comparing reconstructed dMRI tract profiles to the original input, Detect identifies where the deviations occur along white matter pathways. The tool has been successfully applied in identifying microstructural anomalies in various conditions, including epilepsy and schizophrenia. For example, in epilepsy, it detected subtle changes in the white matter beyond areas visible in standard MRI, offering insights into seizure networks. In schizophrenia, Detect identified outliers with higher sensitivity and specificity compared to conventional methods, providing more accurate individual-level diagnosis.

Another significant application of anomaly detection is the work by Kawi et al. [80] which discusses the role of OSVMs in medical formulation recognition (MFR). The authors propose a new Medical Formulation Engine that enables intelligent searches across pharmaceutical archives to enhance the efficiency of formulation recognition and development. It faces the challenge of distinguishing between formulations and non-formulation texts as an anomaly detection problem. In this context, pharmaceutical formulations are considered the “normal” class, while non-formulation texts are treated as anomalies. To address this, the study employs OSVM, highlighting the suitability of the model for identifying these rare instances within large, unbalanced datasets. The proposed MFR system achieved promising accuracy, with a mean recognition rate of 75.2% across experiments and a maximum accuracy of 82%. There are also examples of how this type of UL is applied to images [81]. VasNet, a vasculature-aware UL framework, is specifically designed to address the challenges posed by complex medical imaging data. It uses domain adversarial neural networks (DANNs) to identify and extract relevant vascular structures from noisy, unlabeled imaging data. VasNet distinguishes confounding elements, such as background noise and scattering, from critical features of the vasculature. It then reconstructs vascular images with high fidelity, enabling the detection of structural abnormalities such as thrombosis, internal hemorrhage, and other vascular pathologies.

6. Conclusions

In conclusion, the integration of UL into PM has the power to improve healthcare by analyzing vast and complex biological data. As life sciences and high-throughput technologies advance rapidly, UL offers a scalable framework for extracting valuable insights from heterogeneous and multidimensional data, which often defy interpretation using traditional approaches. Techniques such as clustering, dimensionality reduction, and anomaly detection enable the identification of novel disease subtypes, biomarkers, and patient-specific treatment responses, thereby facilitating the development of more tailored therapeutic strategies.

A key advantage of UL is its ability to uncover hidden patterns in the data without prior assumptions or labels, making it particularly useful for discovering unknown relationships within complex biological systems. This capacity is crucial for identifying previously unrecognized disease mechanisms or patient phenotypes that may lead to new therapeutic targets and precision interventions. While challenges such as data privacy and the need for rigorous validation persist, the ongoing collaboration among scientists remains

essential for fully unlocking UL's potential in PM. It should not be underestimated that integrating AI into PM also involves important ethical considerations. Addressing data privacy, mitigating bias, and ensuring transparency of AI models is critical to maintaining patient trust and achieving equitable health outcomes. The complexity of AI systems requires the development of clear guidelines for healthcare professionals to ensure that these technologies are used responsibly. Looking ahead, it is essential to endorse interdisciplinary collaboration among data scientists, clinicians, and ethicists to ensure the responsible and equitable application of UL in healthcare. Clinicians provide domain knowledge, while data scientists and bioinformaticians offer the technical know-how needed to develop and refine algorithms. At the same time, ethicists are essential for guiding the responsible use of AI, ensuring transparency, mitigating bias, and ensuring patient trust. Future efforts should also focus on creating standardized frameworks for the ethical deployment of UL in PM, particularly with respect to data privacy and model interpretability. As this integration continues to evolve, UL will not only contribute to improving patient outcomes but will also expand the boundaries of disease prevention, diagnostics, and treatment by tailoring care to the individual molecular profiles and life circumstances of each patient. This personalized approach marks a significant leap forward in delivering more effective, efficient, and equitable healthcare for all.

Author Contributions: Conceptualization, A.T., A.V., and B.R.; methodology, A.T., A.V., and B.R.; investigation, A.T., A.V., and B.R.; resources, A.T., A.V., and B.R.; data curation, A.T., A.V., and B.R.; writing—original draft preparation, A.T., A.V., and B.R.; writing—review and editing, A.S. and O.S.; supervision, A.S. and O.S.; project administration, A.S.; funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Johnson, K.B.; Wei, W.Q.; Weeraratne, D.; Frisse, M.E.; Misulis, K.; Rhee, K.; Zhao, J.; Snowdon, J.L. Precision Medicine, AI, and the Future of Personalized Health Care. *Clin. Transl. Sci.* **2021**, *14*, 86–93. [[CrossRef](#)] [[PubMed](#)]
2. Elsevier. *Attitudes to AI: A Global Survey of AI Use and Trust in Healthcare*; Elsevier: Amsterdam, The Netherlands, 2024.
3. Babbage, C. *On the Application of Machinery to the Computation of Astronomical and Mathematical Tables*; Taylor: London, UK, 1824.
4. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 4th ed.; Pearson: Hoboken, NJ, USA, 2021; ISBN 978-0134610993.
5. Bengio, Y.; Courville, A.; Goodfellow, I. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
6. Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. *J. Phys. Conf. Ser.* **2018**, *1142*, 012012. [[CrossRef](#)]
7. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006.
8. Hastie, T.; Tibshirani, R.; Friedman, J.H.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
9. Cunningham, P.; Cord, M.; Delany, S.J. Supervised Learning. In *Machine Learning Techniques for Multimedia*; Cord, M., Cunningham, P., Eds.; Cognitive Technologies: Austin, TX, USA; Springer: Berlin/Heidelberg, Germany, 2008; pp. 21–49.
10. Valkenburg, D.; Rousseau, A.-J.; Geubbelmans, M.; Burzykowski, T. Unsupervised learning. *Am. J. Orthod. Dentofac. Orthop.* **2023**, *163*, 877–882. [[CrossRef](#)]
11. Eckhardt, C.M.; Madjarova, S.J.; Williams, R.J.; Ollivier, M.; Karlsson, J.; Pareek, A.; Nwachukwu, B.U. Unsupervised machine learning methods and emerging applications in healthcare. *Knee Surg. Sports Traumatol. Arthrosc.* **2023**, *31*, 376–381. [[CrossRef](#)]
12. Saxena, A.K.; Prasad, M.; Gupta, A.; Bharill, N.; Patel, O.P.; Tiwari, A.; Er, M.J.; Ding, W.; Lin, C.-T. A review of clustering techniques and developments. *Neurocomputing* **2017**, *267*, 664–681. [[CrossRef](#)]
13. Karim, M.R.; Beyan, O.; Zappa, A.; Costa, I.G.; Rebholz-Schuhmann, D.; Cochez, M.; Decker, S. Deep learning-based clustering approaches for bioinformatics. *Brief. Bioinform.* **2021**, *22*, 393–415. [[CrossRef](#)] [[PubMed](#)]
14. Khan, A.R.; Khan, S.; Harouni, M.; Abbasi, R.; Iqbal, S.; Mehmood, Z. Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification. *Microsc. Res. Tech.* **2021**, *84*, 1389–1399. [[CrossRef](#)] [[PubMed](#)]
15. Jia, W.; Sun, M.; Lian, J.; Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **2022**, *8*, 2663–2693. [[CrossRef](#)]

16. Nassif, A.B.; Talib, M.A.; Nasir, Q.; Dakalbab, F.M. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access* **2021**, *9*, 78658–78700. [[CrossRef](#)]
17. Kascenas, A.; Pugeault, N.; O’Neil, A.Q. Denoising Autoencoders for Unsupervised Anomaly Detection in Brain MRI. In Proceedings of the International Conference on Medical Imaging with Deep Learning, Zurich, Switzerland, 6–8 July 2022.
18. Montavon, G.; Kauffmann, J.; Samek, W.; Müller, K.R. Explaining the Predictions of Unsupervised Learning Models. In *xxAI—Beyond Explainable AI*; Holzinger, A., Goebel, R., Fong, R., Moon, T., Müller, K.R., Samek, W., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2022; Volume 13200, pp. 159–177.
19. Selim, S.Z.; Ismail, M.A. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 81–87. [[CrossRef](#)]
20. Maji, A.; Velaga, N.R.; Urie, Y. Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies. *Int. J. Inj. Control Saf. Promot.* **2018**, *25*, 257–271. [[CrossRef](#)] [[PubMed](#)]
21. Ikotun, A.M.; Ezugwu, A.E.; Abualigah, L.; Abuhaija, B.; Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* **2023**, *622*, 178–210. [[CrossRef](#)]
22. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374*, 20150202. [[CrossRef](#)]
23. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
24. Bank, D.; Koenigstein, N.; Giryas, R. Autoencoders. *arXiv* **2020**, arXiv:2003.05991.
25. Alam, S.; Sonbhadra, S.; Agarwal, S.; Nagabhushan, P. One-class support vector classifiers: A survey. *Knowl.-Based Syst.* **2020**, *196*, 105754. [[CrossRef](#)]
26. Cheng, Z.; Wang, S.; Zhang, P.; Wang, S.; Liu, X.; Zhu, E. Improved autoencoder for unsupervised anomaly detection. *Int. J. Intell. Syst.* **2021**, *36*, 7103–7125. [[CrossRef](#)]
27. Muhtarremi, G.; Meçani, R.; Muka, T. The Buzz Surrounding Precision Medicine: The Imperative of Incorporating It into Evidence-Based Medical Practice. *J. Pers. Med.* **2023**, *14*, 53. [[CrossRef](#)]
28. Cecchin, E.; Stocco, G. Pharmacogenomics and Personalized Medicine. *Genes* **2020**, *11*, 679. [[CrossRef](#)] [[PubMed](#)]
29. Visibelli, A.; Cicaloni, V.; Spiga, O.; Santucci, A. Computational Approaches Integrated in a Digital Ecosystem Platform for a Rare Disease. *Front. Mol. Med.* **2022**, *2*, 827340. [[CrossRef](#)]
30. Marques, L.; Costa, B.; Pereira, M.; Silva, A.; Santos, J.; Saldanha, L.; Silva, I.; Magalhães, P.; Schmidt, S.; Vale, N. Advancing Precision Medicine: A Review of Innovative In Silico Approaches for Drug Development, Clinical Pharmacology and Personalized Healthcare. *Pharmaceutics* **2024**, *16*, 332. [[CrossRef](#)]
31. Visibelli, A.; Roncaglia, B.; Spiga, O.; Santucci, A. The Impact of Artificial Intelligence in the Odyssey of Rare Diseases. *Biomedicines* **2023**, *11*, 887. [[CrossRef](#)]
32. Moeini, R.; Memariani, Z.; Pasalar, P.; Gorji, N. Historical Root of Precision Medicine: An Ancient Concept Concordant with the Modern Pharmacotherapy. *Daru* **2017**, *25*, 7. [[CrossRef](#)]
33. European Commission. Directorate-General for Health and Consumers, Unit D3 eHealth and Health Technology Assessment. In *The Use of Big Data in Public Health Policy and Research*; European Commission: Brussels, Belgium, 2014.
34. Maier, M. Personalized Medicine—A Tradition in General Practice! *Eur. J. Gen. Pract.* **2019**, *25*, 63–64. [[CrossRef](#)]
35. Naithani, N.; Sinha, S.; Misra, P.; Vasudevan, B.; Sahu, R. Precision Medicine: Concept and Tools. *Med. J. Armed Forces India* **2021**, *77*, 249–257. [[CrossRef](#)]
36. König, I.R.; Fuchs, O.; Hansen, G.; von Mutius, E.; Kopp, M.V. What Is Precision Medicine? *Eur. Respir. J.* **2017**, *50*, 1700391. [[CrossRef](#)]
37. Carrasco-Ramiro, F.; Peiró-Pastor, R.; Aguado, B. Human Genomics Projects and Precision Medicine. *Gene Ther.* **2017**, *24*, 551–561. [[CrossRef](#)]
38. Duan, X.P.; Qin, B.D.; Jiao, X.D.; Liu, K.; Wang, Z.; Zang, Y.-S. New Clinical Trial Design in Precision Medicine: Discovery, Development and Direction. *Sig. Transduct. Target. Ther.* **2024**, *9*, 57. [[CrossRef](#)]
39. Fokunang, C.N.; Ndikum, V.; Tabi, O.Y.; Jiofack, R.B.; Ngameni, B.; Guedje, N.M.; Tembe-Fokunang, E.A.; Tomkins, P.; Barkwan, S.; Kechia, F.; et al. Traditional Medicine: Past, Present and Future Research and Development Prospects and Integration in the National Health System of Cameroon. *Afr. J. Tradit. Complement. Altern. Med.* **2011**, *8*, 284–295. [[CrossRef](#)]
40. Li, X.L.; Zhang, J.Q.; Shen, X.J.; Zhang, Y.; Guo, D.A. Overview and Limitations of Database in Global Traditional Medicines: A Narrative Review. *Acta Pharmacol. Sin.* **2024**. [[CrossRef](#)]
41. Mathur, S.; Sutton, J. Personalized Medicine Could Transform Healthcare. *Biomed Rep.* **2017**, *7*, 3–5. [[CrossRef](#)]
42. Baccarelli, A.; Dolinoy, D.C.; Walker, C.L. A Precision Environmental Health Approach to Prevention of Human Disease. *Nat. Commun.* **2023**, *14*, 2449. [[CrossRef](#)]
43. Velmovitsky, P.E.; Bevilacqua, T.; Alencar, P.; Cowan, D.; Morita, P.P. Convergence of Precision Medicine and Public Health into Precision Public Health: Toward a Big Data Perspective. *Front. Public Health* **2021**, *9*, 561873. [[CrossRef](#)]
44. Hasanzad, M.; Sarhangi, N.; Ehsani Chimeh, S.; Ayati, N.; Afzali, M.; Khatami, F.; Nikfar, S.; Meybodi, H.R.A. Precision Medicine Journey through Omics Approach. *J. Diabetes Metab. Disord.* **2021**, *21*, 881–888. [[CrossRef](#)]
45. Tebani, A.; Afonso, C.; Marret, S.; Bekri, S. Omics-Based Strategies in Precision Medicine: Toward a Paradigm Shift in Inborn Errors of Metabolism Investigations. *Int. J. Mol. Sci.* **2016**, *17*, 1555. [[CrossRef](#)]

46. Antonatos, C.; Asmenoudi, P.; Panoutsopoulou, M.; Vasilopoulos, Y. Pharmac-Omics in Psoriasis: Paving the Way towards Personalized Medicine. *Int. J. Mol. Sci.* **2023**, *24*, 7090. [[CrossRef](#)]
47. Liu, Y.; Song, F.; Li, Z.; Chen, L.; Xu, Y.; Sun, H.; Chang, Y. A Comprehensive Tool for Tumor Precision Medicine with Pharmac-Omics Data Analysis. *Front. Pharmacol.* **2023**, *14*, 1085765. [[CrossRef](#)]
48. Sneha, P.; Doss, C.G.P. Molecular Dynamics: New Frontier in Personalized Medicine. *Adv. Protein Chem. Struct. Biol.* **2016**, *102*, 181–224.
49. Trezza, A.; Spiga, O.; Mugnai, P.; Saponara, S.; Sgaragli, G.; Fusi, F. Functional, Electrophysiology, and Molecular Dynamics Analysis of Quercetin-Induced Contraction of Rat Vascular Musculature. *Eur. J. Pharmacol.* **2022**, *918*, 174778. [[CrossRef](#)]
50. Hulsen, T.; Jamuar, S.S.; Moody, A.R.; Karnes, J.H.; Varga, O.; Hedensted, S.; Spreafico, R.; Hafler, D.A.; McKinney, E.F. From Big Data to Precision Medicine. *Front. Med.* **2019**, *6*, 34. [[CrossRef](#)]
51. Prospero, M.; Min, J.S.; Bian, J.; Modave, F. Big Data Hurdles in Precision Medicine and Precision Public Health. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 139. [[CrossRef](#)]
52. Visibelli, A.; Peruzzi, L.; Poli, P.; Scocca, A.; Carnevale, S.; Spiga, O.; Santucci, A. Supporting Machine Learning Model in the Treatment of Chronic Pain. *Biomedicines* **2023**, *11*, 1776. [[CrossRef](#)]
53. Frusciante, L.; Visibelli, A.; Geminiani, M.; Santucci, A.; Spiga, O. Artificial Intelligence Approaches in Drug Discovery: Towards the Laboratory of the Future. *Curr. Top. Med. Chem.* **2022**, *22*, 2176–2189. [[CrossRef](#)]
54. Schork, N.J. Artificial Intelligence and Personalized Medicine. *Cancer Treat. Res.* **2019**, *178*, 265–283.
55. Farhud, D.D.; Zokaei, S. Ethical Issues of Artificial Intelligence in Medicine and Healthcare. *Iran. J. Public Health* **2021**, *50*, i–v. [[CrossRef](#)]
56. Yadav, N.; Pandey, S.; Gupta, A.; Dudani, P.; Gupta, S.; Rangarajan, K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. *Indian Dermatol. Online J.* **2023**, *14*, 788–792. [[CrossRef](#)]
57. McMahan, H.B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. *PMLR* **2016**, *54*, 1273–1282.
58. Teo, Z.L.; Jin, L.; Li, S.; Miao, D.; Zhang, X.; Ng, W.Y.; Tan, T.F.; Lee, D.M.; Chua, K.J.; Heng, J.; et al. Federated Machine Learning in Healthcare: A Systematic Review on Clinical Applications and Technical Architecture. *Cell Rep. Med.* **2024**, *5*, 101419. [[CrossRef](#)]
59. Norori, N.; Hu, Q.; Aellen, F.M.; Faraci, F.D.; Tzovara, A. Addressing Bias in Big Data and AI for Health Care: A Call for Open Science. *Patterns* **2021**, *2*, 100347. [[CrossRef](#)]
60. Liu, M.; Ning, Y.; Ke, Y.; Shang, Y.; Chakraborty, B.; Ong, M.E.H.; Vaughan, R.; Liu, N. FAIM: Fairness-Aware Interpretable Modeling for Trustworthy Machine Learning in Healthcare. *Patterns* **2024**. [[CrossRef](#)]
61. Jones, C.; Thornton, J.; Wyatt, J.C. Artificial Intelligence and Clinical Decision Support: Clinicians’ Perspectives on Trust, Trustworthiness, and Liability. *Med. Law Rev.* **2023**, *31*, 501–520. [[CrossRef](#)]
62. Robinson, G.A.; Peng, J.; Dönnnes, P.; Coelewijn, L.; Naja, M.; Radziszewska, A.; Wincup, C.; Peckham, H.; A Isenberg, D.; Ioannou, Y.; et al. Disease-Associated and Patient-Specific Immune Cell Signatures in Juvenile-Onset Systemic Lupus Erythematosus: Patient Stratification Using a Machine-Learning Approach. *Lancet Rheumatol.* **2020**, *2*, e485–e496. [[CrossRef](#)]
63. Nelke, C.; Schroeter, C.B.; Barman, S.; Stascheit, F.; Masanneck, L.; Theissen, L.; Huntemann, N.; Walli, S.; Cengiz, D.; Dobelmann, V.; et al. Identification of Disease Phenotypes in Acetylcholine Receptor-Antibody Myasthenia Gravis Using Proteomics-Based Consensus Clustering. *eBioMedicine* **2024**, *105*, 105231. [[CrossRef](#)]
64. Smith, J.; Lee, K.; Patel, R. Precision Medicine in Diabetes: Application of K-Means Clustering to Optimize Treatment. *J. Med. Inform.* **2024**, *12*, 105–120.
65. Zhang, X.; Li, Y.; Wu, H.; Guo, S. Using K-Means Clustering to Identify Novel Phenotypes of Acute Ischemic Stroke. *Front. Neurol.* **2024**, *15*, 224.
66. Dutta, A.; Pal, A.; Bhadra, M.; Khan, M.A.; Chakraborty, R. An Improved K-Means Algorithm for Effective Medical Image Segmentation. *Math. Comput. Model.* **2024**, *in press*.
67. Madhukar, N.S.; Khade, P.K.; Huang, L.; Gayvert, K.; Galletti, G.; Stogniew, M.; Allen, J.E.; Giannakakou, P.; Elemento, O. A Bayesian Machine Learning Approach for Drug Target Identification Using Diverse Data Types. *Nat. Commun.* **2019**, *10*, 5221. [[CrossRef](#)]
68. Gerdes, H.; Casado, P.; Dokal, A.; Hijazi, M.; Akhtar, N.; Osuntola, R.; Rajeeve, V.; Fitzgibbon, J.; Travers, J.; Britton, D.; et al. Drug Ranking Using Machine Learning Systematically Predicts the Efficacy of Anti-Cancer Drugs. *Nat. Commun.* **2021**, *12*, 1850. [[CrossRef](#)]
69. Bonazzola, R.; Ferrante, E.; Ravikumar, N.; Xia, Y.; Keavney, B.; Plein, S.; Syeda-Mahmood, T.; Frangi, A.F. Unsupervised Ensemble-Based Phenotyping Enhances Discoverability of Genes Related to Left-Ventricular Morphology. *Nat. Mach. Intell.* **2024**, *6*, 291–306. [[CrossRef](#)]
70. Leonardsen, E.H.; Persson, K.; Grødem, E.; Dinsdale, N.; Schellhorn, T.; Roe, J.M.; Vidal-Piñeiro, D.; Sørensen, Ø.; Kaufmann, T.; Westman, E.; et al. Constructing Personalized Characterizations of Structural Brain Aberrations in Patients with Dementia Using Explainable Artificial Intelligence. *NPJ Digit. Med.* **2024**, *7*, 110. [[CrossRef](#)]
71. Sturm, D.; Capper, D.; Andreiuolo, F.; Gessi, M.; Kölsche, C.; Reinhardt, A.; Sievers, P.; Wefers, A.K.; Ebrahimi, A.; Suwala, A.K.; et al. Multiomic Neuropathology Improves Diagnostic Accuracy in Pediatric Neuro-Oncology. *Nat. Med.* **2023**, *29*, 917–926. [[CrossRef](#)] [[PubMed](#)]

72. Kropivsek, K.; Kachel, P.; Goetze, S.; Wegmann, R.; Festl, Y.; Severin, Y.; Hale, B.D.; Mena, J.; van Drogen, A.; Dietliker, N.; et al. Ex Vivo Drug Response Heterogeneity Reveals Personalized Therapeutic Strategies for Patients with Multiple Myeloma. *Nat. Cancer* **2023**, *4*, 734–753. [[CrossRef](#)] [[PubMed](#)]
73. Lin, J.R.; Wang, S.; Coy, S.; Chen, Y.-A.; Yapp, C.; Tyler, M.; Nariya, M.K.; Heiser, C.N.; Lau, K.S.; Santagata, S.; et al. Multiplexed 3D Atlas of State Transitions and Immune Interaction in Colorectal Cancer. *Cell* **2023**, *186*, 363–381.e19. [[CrossRef](#)]
74. Mahdavi, M.; Choubdar, H.; Rostami, Z.; Niroomand, B.; Levine, A.T.; Fatemi, A.; Bolhasani, E.; Vahabie, A.-H.; Lomber, S.G.; Merrikhi, Y. Hybrid Feature Engineering of Medical Data via Variational Autoencoders with Triplet Loss: A COVID-19 Prognosis Study. *Sci. Rep.* **2023**, *13*, 2827. [[CrossRef](#)] [[PubMed](#)]
75. Yuan, B.; Yang, D.; Rothberg, B.E.G.; Chang, H.; Xu, T. Unsupervised and Supervised Learning with Neural Network for Human Transcriptome Analysis and Cancer Diagnosis. *Sci. Rep.* **2020**, *10*, 19106. [[CrossRef](#)]
76. Wegmann, R.; Bankel, L.; Festl, Y.; Lau, K.; Lee, S.; Arnold, F.; Cappelletti, V.; Fehr, A.; Picotti, P.; Dedes, K.J.; et al. Molecular and Functional Landscape of Malignant Serous Effusions for Precision Oncology. *Nat. Commun.* **2024**, *15*, 8544. [[CrossRef](#)] [[PubMed](#)]
77. Deng, Y.; Yao, Y.; Wang, Y.; Yu, T.; Cai, W.; Zhou, D.; Yin, F.; Liu, W.; Liu, Y.; Xie, C.; et al. An End-to-End Deep Learning Method for Mass Spectrometry Data Analysis to Reveal Disease-Specific Metabolic Profiles. *Nat. Commun.* **2024**, *15*, 7136. [[CrossRef](#)]
78. He, D.; Liu, Q.; Wu, Y.; Xie, L. A Context-Aware Deconfounding Autoencoder for Robust Prediction of Personalized Clinical Drug Response from Cell-Line Compound Screening. *Nat. Mach. Intell.* **2022**, *4*, 879–892. [[CrossRef](#)]
79. Chamberland, M.; Genc, S.; Tax, C.M.W.; Shastin, D.; Koller, K.; Raven, E.P.; Cunningham, A.; Doherty, J.; van den Bree, M.B.; Parker, G.D.; et al. Detecting Microstructural Deviations in Individuals with Deep Diffusion MRI Tractometry. *Nat. Comput. Sci.* **2021**, *1*, 598–606. [[CrossRef](#)]
80. Kawi, O.; Clawson, K.; Dunn, P.; Knight, D.; Hodgson, J.; Peng, Y. Medical Formulation Recognition (MFR) Using Deep Feature Learning and One Class SVM. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
81. Wang, Y.; Ji, M.; Jiang, S.; Wang, X.; Wu, J.; Duan, F.; Fan, J.; Huang, L.; Ma, S.; Fang, L.; et al. Augmenting vascular disease diagnosis by vasculature-aware unsupervised learning. *Nat. Mach. Intell.* **2020**, *2*, 337–346. [[CrossRef](#)]

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.