PMTRS: A Personalized Multimodal Treatment Response System Framework for Personalized Healthcare

Sajib Alam
Trine University

Abstract

Various types of data are used in clinical settings, such as imaging, textual, sequential, and, tabular data. Multimodal machine learning focuses on combining multiple modalities (types) to create an overall representation. This involves extracting features from each modality, combining them into a unified representation, and using these representations to enhance decision-making processes in AI applications. This research introduces the Personalized Multimodal Treatment Response System (PMTRS), a novel framework aimed at enhancing personalized treatment by utilizing multimodal machine learning to analyze various data types, including genetic information, medical imaging, and electronic health records. The proposed PMTRS is designed to predict and optimize individual treatment outcomes through a structured approach comprising several key components. First, the Data Collection and Preprocessing Module is responsible for gathering diverse patient data and preparing it for analysis through normalization and modality-specific processing techniques. The Feature Extraction and Integration Module then applies deep learning models, such as convolutional neural networks for imaging data and natural language processing for electronic health records, to extract relevant features and integrate them using fusion techniques. At the core of PMTRS is the Personalized Treatment Prediction Model, which employs a multimodal deep learning architecture capable of handling integrated features from various data types using supervised learning and incorporates transfer learning to predict treatment responses accurately. The Treatment Recommendation System uses these predictions to provide personalized treatment options, supported by an Explainability Module to ensure transparency and build trust in the system's decisions.

Introduction

The notion of modality typifies the manner in which we approach the world and perceive it. In essence, this term pertains to the different ways in which natural phenomena can be seen or heard, learned or even taught. The concept includes such varied types of perception as aural, oral, tactile and others. For example, microphones collect the sounds of the world, moving from the source to the transceiver. Cameras collect the images, capturing the world on film or digital tape. Haptic sensors even collect the sensations, measuring the texture or stability of any given object. All of these perceptions can be qualified as modalities, representing the diverse means of acquiring information and direct sources of data from the world that surrounds us. As such, there are two types of modalities: raw and abstract. The main difference between these types is the proximity and immediateness to the source, meaning how close these two are to the fact of data collection. In other words, raw modalities are the initial data and direct sensory perception of sensorial input. As such, the examples of these can be sounds the microphone records or an image the camera takes. This type of modality is data in its purest form, recorded as is without any elaboration. However, even at this early stage, such data can be considered a step up from the source as it was transferred across the air from one device to another. Abstract modality defines the state when data collected by sensors is no longer a source and is no longer close to the immediate perceived reality. In abstract modalities, the data is processed in a meaning extracted from speech, recognized objects in the picture, and intensity of sentiment to text. In essence, abstract and raw modalities activate different levels of interpretation of reality.

Table 1. Five core challenges of multimodal learning

Challenge Number	Core Challenge	Description
1	Data Representation and Fusion	Developing methods to effectively represent and integrate diverse data types from different modalities into a unified model for coherent analysis and learning.
2	Heterogeneity of Data	Adapting to and leveraging the unique properties, scales, and types of information presented by different modalities.
3	Alignment and Synchronization	Ensuring that related information from different sources is correctly matched and aligned, particularly in time or space.
4	Missing Data and Imbalance in Modalities	Handling incomplete or unbalanced datasets by developing models that can infer missing information or learn effectively without bias.
5	Scalability and Computational Efficiency	Creating scalable models that can process and analyze large volumes of multimodal data efficiently.

The term “multimodal” is applied to a variety of such cases in which there are multiple ways for modalities to play a role, exists as a provocative concept, and is meant to exemplify the idea that such a variety of options for data and sensory types are at the heart of these. In that sense, when it applies to research, particularly in the property, of computation, and the other sciences, there are types of multimodal research which exist as ones which look into these disparate types and ways of being connected in their forms of relations. Heterogeneity is the characterization of these connections, as the heterogeneous natures of these information differences are what the forms, types, and qualities of these representations can be. Data from each of these ways are put together to be evaluated as to what they may be when experienced as a type of form. It is with these multiple views that the ways can be evaluated with respect to their qualities to be able to view them as the things they are. The multiple forms that the research takes on are important for the study of these connections in a context which offers a greater depth than offered by any one form of modality. Also, these forms connect with each other, as the multiple forms of information are in complementary types of relations to the other forms [5], [6].

Take multimodal deep learning for example, this is an important field of machine learning, and it has a very well worked out methodology for combining data from several sources at once. In the context of a multimodal model, the problem exists to move forward these five issues and provide an elaborate analysis of such complex data in a larger context. The first issue is data representation and fusion. Different sorts of information have to be brought together under one kind of representation, which the computer can use to train an AI model and so ensure that an intelligent explanation is consistent across all modalities. An algorithm must be developed that is capable of processing different kinds of sensory data - the written word, images, spoken language - and bringing them all together in one model. The model that results is a coherent representation of what characteristics in and meaning of the data, exactly like every input. The second issue is the heterogeneity of data. Multiple data types have different properties and scales, making it hard to guarantee that the data of different types can go into one system without any change of its conditions or how often it has samples for each type. Ways need to be found to represent each modality in its own sampling system, accounting for the heterogeneity of the data. The answer to these issues is likely to come at the level of model training, perhaps requiring new types of solutions that can handle data as well as its modality types [7, 8]. The third point, like the second, deals with data alignment. It means that when another piece of related data for an AI model to make a decision whether comes from one of these modalities or their combination, it is time and space–corrected into place surrounding the first [9]. The most direct application of this question is in video, where both audio and video have to be aligned.

The fourth challenge is related to dealing with missing data and the imbalance in modalities. Not all modalities will be available for every instance of data in the real world and as such, datasets can be incomplete or unbalanced. To address this issue, I must develop models and techniques resilient to such imperfections, either capable of gracefully inferring missing information or effectively learning from unbalanced data without bias. The fifth challenge is concerned with scalability and computational efficiency. Since the multimodal data is complex and abundant as discussed above, processing and analyzing it can prove to be resource-intensive. Hence, it is vital for my work to develop scalable models efficiently working with large volumes of multimodal data to enable the practical application of multimodal deep learning in autonomous driving, healthcare, and multimedia analysis.

Architectures of proposed PMTRS

The proposed PMTRS consists of several key components designed to handle different aspects of multimodal data processing, model training, and decision support for personalized treatment.

1. Data Collection and Preprocessing Module

Data Sources: Genetic data (e.g., DNA sequencing), medical imaging (e.g., MRI, CT scans), electronic health records (EHRs), and patient-reported outcomes.

Preprocessing: Data normalization, missing value imputation, and modality-specific processing (e.g., image segmentation for medical imaging).

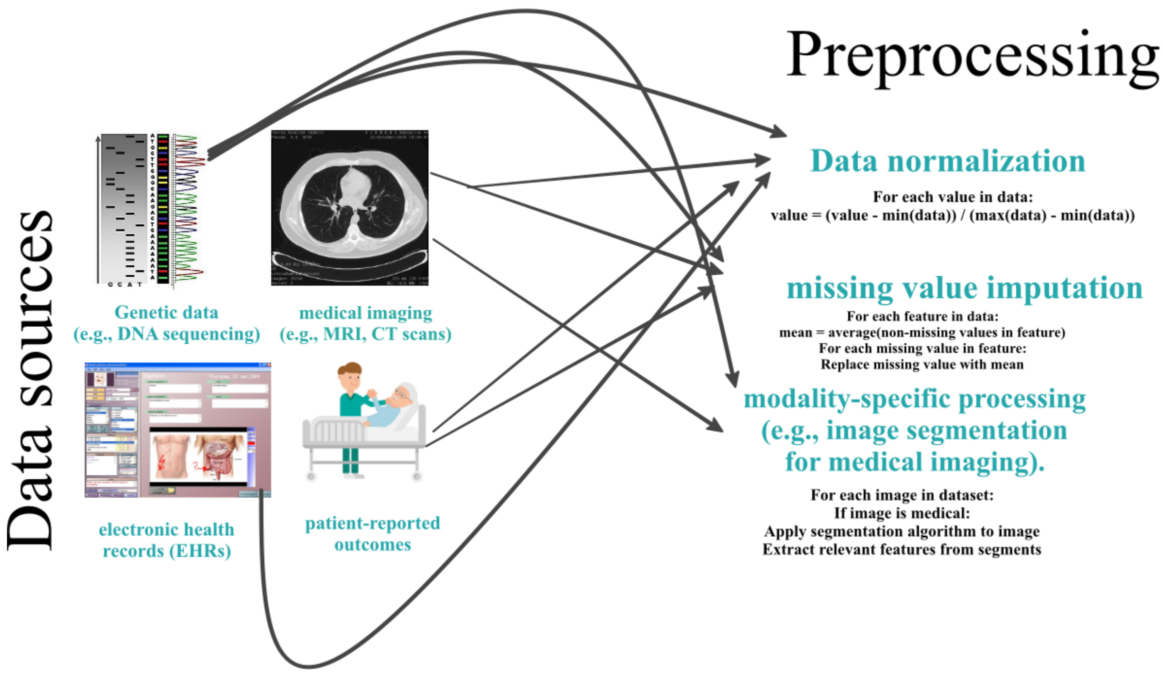

Figure 1. Data Collection and Preprocessing Module

Genetic data, particularly the data obtained through a variety of next generation sequencing methods reveals an extra depth in terms of information available to the patient or any other stakeholder. It provides detailed information on the patients' genetic landscapes including details on such aspects as single nucleotide polymorphisms or SNPs, insertions, deletions, and more complex types of variations. This information is invaluable in several different areas including, though not limited to, the pharmacogenomics and identification of specific markers associated with hereditary diseases such as cystic fibrosis or sickle cell anemia. Genetic data is widely utilized in the context of oncogenomics, a relatively new branch of genomics, aiming to understand cancer at the molecular level. Such an approach facilitates the development of therapies that are specifically designed to affect the genetic abnormalities in the tumours of a particular patient to slow down or stop the growth of the malignant cells.

With the rising importance of genetic data in modern medicine, the advancements in MRI and CT Scan led to higher diagnostic precision and better treatment outcomes. High-resolution MRI, for instance, assists the field of neurology by allowing for early diagnostics of multiple sclerosis and Alzheimers disease. Since both are neurodegenerative diseases that progress at a high speed, early diagnosis is of paramount importance in terms of both prognostics and treatment. Since standard MRI methodologies allow to differentiate between the types of tissues and create an image of the brain, high-resolution MRI adds an extra depth to diagnostics by allowing to detect even minor changes over time and take actions if more dynamic changes are detected. The CT scans, in turn, perform a similar function but are powered by an X-ray instead of magnetic energy. They are much more widely used in the emergency and trauma setting to provide a quick, detailed, and high-resolution assessment of the patients condition. The time between patient admission to a hospital and the diagnostic procedure is particularly important in trauma patients, as bleeding or other conditions causing critical conditions and in need of a surgery can be detected within minutes or even seconds of a arrival.

Both Electronic Health Records that capture a patient's data beyond the point of care and over time and patient-reported outcomes capturing the patient's perspective on their health or their treatment meaningfully converge in a number of aspects in healthcare. On the one hand, EHRs provide a way to take a "big picture" view of a patient's health not only in terms of personal data and administration but in diagnosing and providing healthcare options. EHRs integration thus adds an important link to clinical applications to take data in order to make it more useful. On the other hand, patient-reported outcomes are often gathered in surveys expanding the data on patients' health treatments to encompass the patient's view of treatment effectiveness, relief of their symptoms, how it affects their daily living, and overall satisfaction with services. Their aggregation subsequently enhances decision-making for quality of care from the patients' perspective. Their fusion can provide a better perspective on diagnostic and treatment outcomes and therefore can likely improve healthcare.

Data normalization is a vital step in data preprocessing prior to data analysis as it translates and adjusts different variables based on the measurements to a different common scale. The main rationale for data normalization is improved comparability and effectiveness of algorithms. It is also particularly relevant for datasets like genetic data that consist of gene expressions measured across multiple tissues that are produced in different laboratories and conditions and thus constitute data of varying dimensions and scales. Missing value imputation is the process of populating datasets that contain many default, null, or missing numbers. Both techniques are invaluable for neurologists and healthcare professionals, where data is often flawed, and missing data are the least of their concerns. In the case of missing value imputation, classical and advanced techniques can fill those gaps and make the datasets complete and ready for processing.

When it comes to medical imaging, that which occurs within an image specifically for a particular modality is key in identifying accurately and defining structures and regions of interest. The purpose of segmentation is mainly to divide medical image into segments, which may either simplify itself or change it into something meaningful. Thus tumor segments can be generated from a scan so that the growth is isolated from surrounding tissue, allowing for exact space and shape measurements as well as change in development over time. This type of modality-specific processing is the core of medical imaging rendering it indispensable for clinical diagnosis, treatment and care monitoring.

2. Feature Extraction and Integration Module

Feature Extraction: Use deep learning models to pull key features from each modality of data. For example, imaging data by convolutional neural networks (CNNs) and EHR by natural language processing (NLP) methods.

Integration Techniques: Use fusion methods (early fusion, late fusion, model-based integration)to combine the features from different modalities well, so that the model benefits from what is complementary between each kind of data.

Data normalization, a fundamental preprocessing step, ensures that diverse data sets,

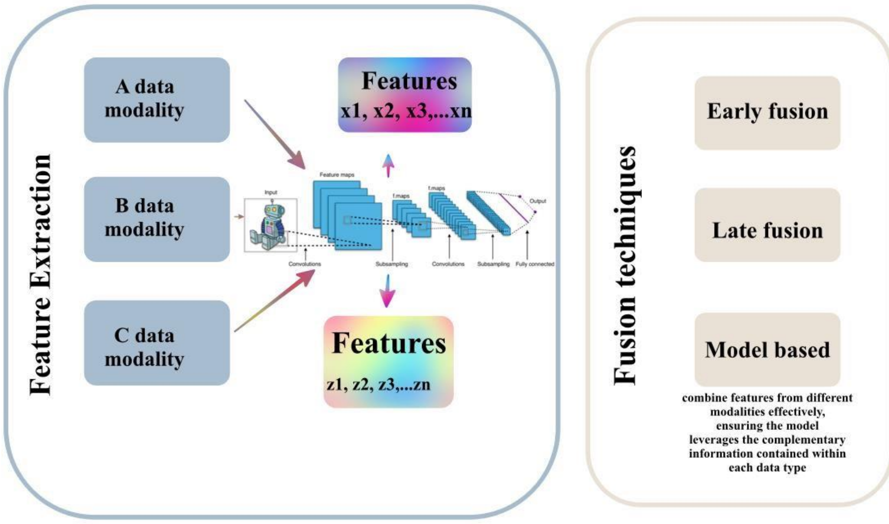

Figure 2. Feature Extraction and Integration Module

The first step is data normalization, which is used when the information and signals, particularly those derived from heterogeneous sources, are brought to a common scale without distorting differences in the ranges of values. This process is instrumental in genetic data analysis, meaning that gene expression levels from different patients need to be compared. To this end, such techniques as z-score normalization or min-max scaling are applied to adjust the data. As for missing value imputation, it helps to address the widespread problem of incomplete datasets found in many areas of medical research, in EHRs and genetic data in particular. In terms of the above processes, a number of statistical and machine learning methods are used, including k-nearest neighbors imputation, expectation maximization , and multiple imputation by chained equations. According to Coşkun et al., missing value imputation is used to fill these gaps and avoid biased decisions or models when the information and data is not complete. Lastly, as far as modality-specific processing is concerned, it is also associated with the aforementioned area and such issues as medical imaging, in particular. For instance, one of the used techniques involves image segmentation which is critical when identifying specific structures or regions of interest within MRI or CT scans. In general, the used algorithms, such as thresholding, region-growing, and a number of machine learning-based methods, help to isolate anatomical regions or pathological features from the "background" tissue. This precision is vital when quantifying tumor volumes, evaluating disease progression, or planning surgeries where the actual tissues involved should be accounted for.

Many deep learning models perform well with new data in different complex forms.There are numerous models can be used to implement deep learning in various data modality domains Images are represented in data format, bigger image data are more difficult to process, and where storage capacity is at a premium compared to the more manageable amounts taken by each individual character on a typical screen. The difference between it Footballs shaped a certain way and flying pig balloons is that the actual connotation for each example the first football while metaphorically the second is well-knownIndeed, CNNs are structured in such a way that they tune themselves to fit this kind of picture of imaging data, processing each level in the hierarchy as a cluster features or other combinations of characteristics along with every feature comprising part. In medical and health applications, this kind of model can be used to decide whether there are patterns at all e.g. in MRI or CT scans tumor or rather just some kind of fracture. Hence, when dealing with image data, CNNs have an advantage over other processing techniques in terms of efficiency. Electronic health records, among other text-based data types, pose the problem of extracting relevant information or relationships between different blocks of data. Here, it is techniques of natural language processing tailored first and foremost for free text that come to the rescue. Tokenization, PoS tagging, and NER are some of the methods used to change unstructured data into a structured format that is convenient for analysis.

Some content is different when it comes to imaging versus free text data. Even so, they can finally bring forth much nourishment in a great variety of health care analysis topics: predictive analytics and health-related decisions for instance. Integration techniques mostly refer to different types of data, like genetic sequences and MRI or EHR data. Because there is sure to be some deviation between the two, all possible discrepancies must be tracked down and eliminated to form a more complete result. Multimodal learning is the early fusion or feature-level integration method. It combines different kinds of information at an early stage, before training a model, so that the system can then work with a single feature vector. This was chosen as the first choice because combining modalities allows the model to learn from the richly diverse dataset. It can exploit interrelations learned between different data types and their significance within records. For instance, one can combine molecular genetic information with imaging data to make a feature-rich vector which describes a disease in both its molecular nature and phenotypic appearances. In this case the model would be expected to have better performance when predicting the progression of a disease, its response to therapy, or other characteristics which are relevant to the problem at hand.

As a second way of comprehending, this late coupling process- also known as decision-level integration involves decision level aggregations of independently developed models operating on different kinds of information. As long as the order of the techniques is so chosen, selecting this as the second best approach may result in a fusion of results from a rich set of applying independently working models to, for example, MRI scans and EHR narratives. At the same time, import information about imaging resolution and scale could be complemented with the temporal information from EHRs so that the precision of the data set is raised. Furthermore, decision-level integration means that separate models conduct their data analysis through this kind of process and order more model flavors are discarded, which is highly efficient and friendly to hardware. There are many forms of data fusion techniques. These techniques enable clinical researchers to develop model that integrate information from multiple sources while giving an overall understanding of oral health, which offers the prospect for more accurate diagnosis, prognosis and treatment. The choice of fusion technique depends on the nature of data being handled as well as what exact clinical question is posed: people vary their model complexity and interpretability according to this grasp.

3. Personalized Treatment Prediction Model

Modeling Structure: A multimodal deep learning model capable of incorporating the integrated features from different forms of data. This could involve a blend of deep learning algorithms such as CNNs for image data and recurrent neural networks (RNNs) in connection with genetic data; fully connected layers could also be used for genomic information. Each of these subsystems would train on its own.Markov Chain Monte Carlo However, it is not easy to summarized in a couple of simple sentences.

Learning Approach: Through supervised learning to predict treatment responses given historical treatment results. It also involves techniques like transfer learning which translates knowledge from other datasets in the same domain, and for the purpose of improving generalization ability.

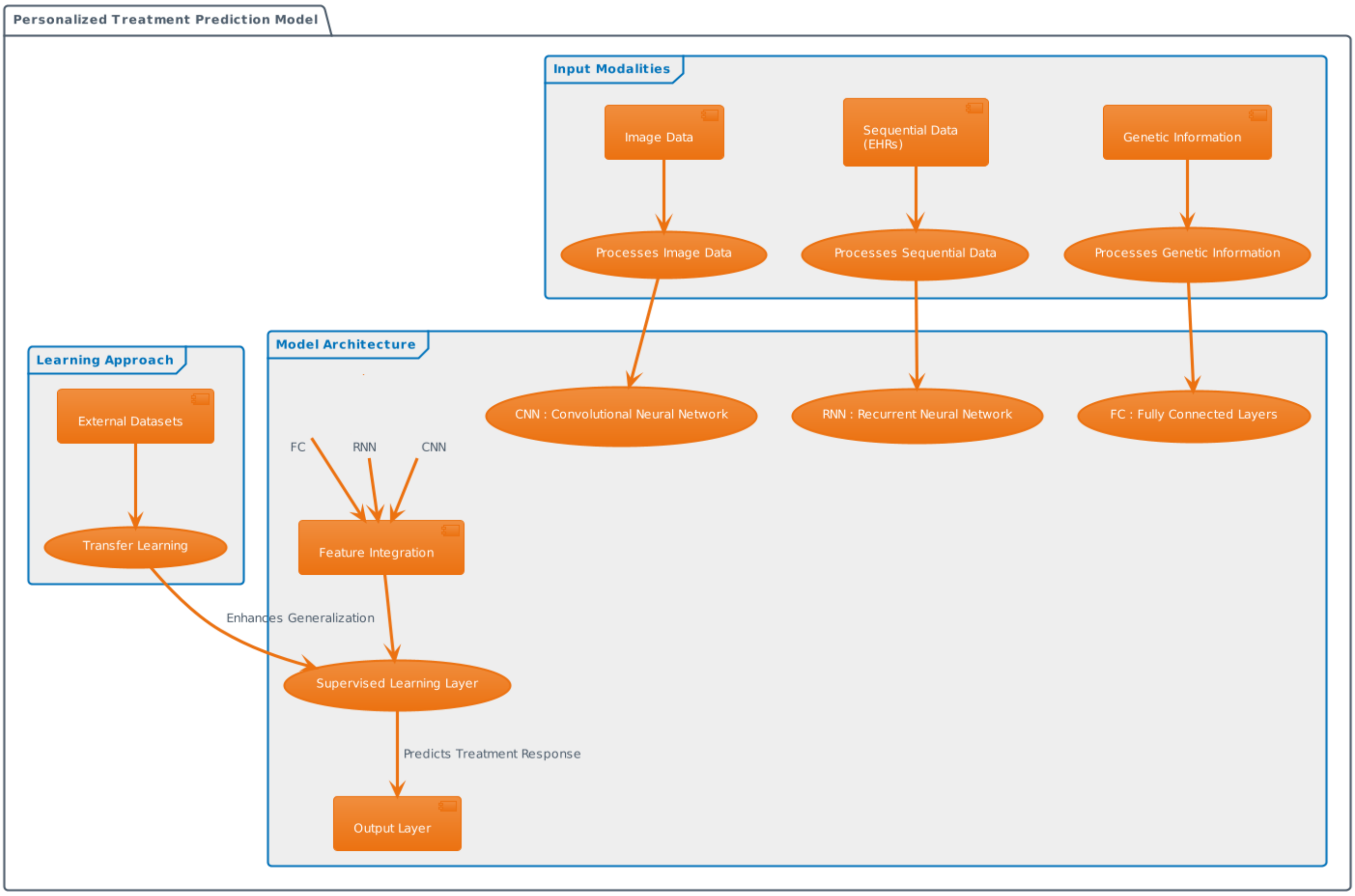

Figure 3. Personalized Treatment Prediction Model

The architecture of such a model for predicting the results of a personalized treatment in the field of healthcare is based on the advantages of multimodal deep learning, which allows accounting for heterogeneous data types. Each model incorporated into the architecture is specifically developed for analysis of different data types. CNNs are used for the analysis of images, with medical images passed through the "layers" of the model. This part of the architecture recognizes patterns in images of organs, body parts, and various objects, taking into consideration different stages of the effectiveness of these methods. At the same time, RNNs analyze EHRs through learning how to recognize the scheduled changes, meaning to understand what exactly and where happens that may range from the testing of changes to determine the frequency of some types of intervention. Lastly, genetic sequences are analyzed through standard fully-connected layers analyzing multi-dimensionality to identify how gene types and amount of mutations affect the response to intervention. The combination of these models is used to ensure that the entire picture of the patient, on the basis of which the predictions will be conducted, is taken into consideration in each model, meaning that the unique data that serves as the source of information for different modalities will be used and will be considered in the models' holistic approach to the prediction.

The nature of learning of such a predictive model is supervised learning, where the model is exposed to a dataset and is left to predict the responses of various patients to the intervention. The dataset of the model is related to the pre-collected data on patients and the data on the intervention, including images, genetic sequences, and EHR data. In addition, the data on interventions is included in the dataset in the form of different labels marking the response to such an intervention and the result. The use of this approach by the model is characterized by the need to thoroughly understand the relationships between multimodal features and the likelihood of the response to the intervention phases of the individual on them. In addition, the approach to learning is integrated with transfer learning, which is used as the answer to the problem of the generalization of the model and its adaptation to new patient populations and datasets. In the light of this approach, the model is previously trained on vast data to develop features and patterns included in the final model's generalization for future learning and functioning on more narrow or specific datasets.

4. Treatment Recommendation System

- **Decision Support:** Utilize the predictive model to generate personalized treatment recommendations, highlighting the predicted effectiveness of various treatment options for the specific patient.

Explainability Module: Incorporating the use of explainable AI (XAI) techniques can enlighten the model's decision process and foster greater trust within healthcare professionals and patients towards these artificial intelligence tools.

The personalized treatment prediction model in which the Treatment Recommendation System operates is at the crossroads of advanced predictive modeling and decision support systems. This system painstakingly analyzes how predicted responses will look when a patient is treated by various means - every piece of multimodal patient data from genetic information, medical imaging, and electronic health records is brought together. Based on the probability of different treatments succeeding, the system instructs medical personnel on what therapeutic approach to take for each patient, thereby promoting more efficacious treatment outcomes. And because the predictive model can synthesize large volumes of patient data in an all-encompassing way, each proposal it puts forward is steeped in the intricate specific context of the patient's health.

Whether or not this recommendation system is effective depends to a large degree on the incorporation of an Explainability Module.... This module is designed to expose the decision-making process of the model, and thereby provides transparent explanations as to how and why certain treatment suggestions are made. Such things as feature importance visualization, model-agnostic explanation techniques, and counterfactual explanations are key in this context. By demystifying the model's decision process in this way, clinicians and patients can rest easy knowing that the recommendations from each recommendation are not only clear but also trustworthy.

An explainability Module in the Treatment Recommendation System provides a clear rationale for every treatment suggestion. Healthcare professionals can make more informed decisions; patients are able to clearly grasp the various options that exist and take part in their own treatments. This is no mean feat because it means that both parties are better prepared for any situations arising after treatment.Furthermore defaultstate-of-the-art explainability AI techniques applied in the System also constitutes a significantly new departure for personalized medicine. Through the integration of extensive patient data multi-modality with predictive modeling, we are able to both generate and expound upon treatment recommendations. This system looks at a precise approach to patient care in contrast with today's approach: dragging its heels although there is still insufficient data just makes the whole question more complicated. An in-depth grasp of a particular patient's medical history and individual characteristics in this way will lead to more exacting, effective treatment plans.Assuming all goes well, the transparency realized by explainable AI can not only guarantee the system's reliability but also help to produce an all-participating cooperation in therapy: by making patients a part of their own progress.

5. Continuous Learning and Adaptation

- **Feedback Loop:** Incorporate a mechanism for updating the model based on new patient data and outcomes, allowing the system to adapt and improve over time.
- **Privacy and Security:** Implement robust data protection and privacy measures to secure sensitive patient information.

This Feedback Loop mechanism, built specifically for the Continuous Learning and Adaptation platform, is exactly that; it reintroduces nutrients from new patient data to feed back into a system which has learned along its way. Online learning methods, which can have the mechanism to change without re-training an entire model if they use advanced machine learning techniques (e. g., on-line FW and AROW); here once a new piece of information arrives at this utility it updates relevant parameters within its models that depend on it. For instance, if the goal might be to fit weights on a neural network based patient outcomes gathered in practice adjustments depending upon global data being collected and reported via gradients over time. This results in the model becoming more predictable as disease patterns change, and new effective treatments are developed or health care practices vary. Privacy and Security: The system incorporates unique technological functions that safeguard patients' confidentiality. Moreover, all data in transit between the system and care provider networks is encrypted with end-to-end encryption protocols such as TLS (Transport Layer Security), while AES (Advanced Encryption Standard) evncryption ise used to keep at rest stored within system databases. Bridging the gap with high-regulation standards (HIPAA; GDPR): Considering rather stringent regulatory guidelines, a fine-grain access control mechanism over system is implemented through blockchain technology for an auditable trail of information accessed and modified. This not only guarantees that patient data is viewed by the right people, but it also creates a clean audit trail for regulatory purposes. In addition, during training differential privacy should also be employed to increase the protection of patient data which can no longer be re-identified from model outputs or analysis [11]. These individual steps make up a system of patient protection in general and privacy standards for health data to prevent unauthorized use or access.

Implementation Considerations

For the effective implementation of the system of personal treatment prediction and recommendation implementation would require a combined effort that involves multiple fields. The central authority in this endeavor is the synergy between machine learning experts, clinicians, and bioinformaticians. On the one hand, machine learning experts are needed to improve and develop algorithms that would be able to process and analyze the gathered complex, multimodal data. The clinicians are essential to provide input about patient care and the practical implications of the models' predictions. It is also vital for clinicians to oversee the development of the system which recommendations can be used and that would produce ideas that can be transferred into the real world. Lastly, bioinformaticians are necessary to analyze biological data, such as NEC or other genetic information, needed to make the required treatment decisions to personalize. This combined work between the technical work of designing the system and a clinician's work ensures that the created system will not only be state-of-the-art but would also be useful and aligned with the realities of patient care.

Ethical concerns and regulations are the backbone on which the credibility is built.e.g. For instance, the consent provided by the patients should both be explicit, so tells the patients what type of treatment it will allow, and opt-in, meaning that the control over the patient data must be natural and in control of the patient themselves. Within the US and beyond, the medical laws must also be followede.g.. Ethical concerns also pertain to inescapable manner the question of how to ethically develop a system that would not inadvertently create models that favor treatment types for the disadvantaged patients. The initiative and the latter goals have led the project to ensure that ethical considerations and regulatory compliance are implemented within every step of the way.

Finally, rigorous clinical validation is also necessary to ensure that the implemented system is both beneficial and safe. It could be conducted through clinical trials that would compare the provided form of treatment with the standard care and analyze patient recoveries as a results. The overall goal is to present empirical evidence that the implemented system can effectively improve the patient outcomes, thus creating a better and more person-specific paradigm of healthcare planning.

Conclusion

The proposed Personalized Multimodal Treatment Response System is comprised of a number of essential modules that can help ensure the accuracy of patient- and medication-specific treatment and intervention. Specifically, the Data Collection and Preprocessing Module is the foundational element of the system, amalgamating data from a variety of sources. The latter include genetic sequences, MRI and CT scans and other types of medical imagery, EHRs, and the patient-reported outcomes. In the context of the developed module, one might need to conduct the process of data normalization, imputation of missing values, and, importantly, the modalityspecific processing that

may include the application of sophisticated tools for image processing and segmentation on medical images. Importantly, the carefully curated content processed with the module can ensure that other modules of the proposed framework operate with the correct, uncontaminated data relevant to the task.
The Feature Extraction and Integration Module is the next step in PMTRS, which utilizes sophisticated deep learning models to narrow down the crucial features of each modality of data. For example, convolutional neural networks are used to analyze the imaging data, and in its turn, natural language processing is employed to extract vital information from the electronic health records. Regarding this module, the principal aspect is its technique of integration. It can be early fusion, late fusion, or model-based integration to combine the features of different modalities. Such techniques are important as they help to parse the supplementary information naturally occurring in each type of data, thus enriching the prediction produced by the model. More specifically, PMTRS introduces the model, capable of processing the integrated feature from multiple data types. In this case, treating options can be combined CNNs are used for the image data, RNNs are processing the sequential data found in EHRs, and the fully connected layer can be used for the genetic data. This model is primarily taught by supervised learning to predict the treatment outcomes by using historical data of the patients. However, transfer learning is often applied to teach the model to generalize by using an additional dataset for validation. Finally, this predictive model is extremely important as it helps to determine the most effective treatment options based on the unique data of each individual. The Treatment Recommendation System is the essential component of PMTRS, as one of the main aspects of the program is providing decision support to the clinicians about the personalized treatment explanation. The system is not purely predictive as it does not only determine the efficacy of different treatment options on a provided patient but also has an Explainability Module. The module is used as the XAI techniques to explain the logic of the model's decisions and justify the provided treatment recommendations. The explainability of the AI decisions, especially in health care, is necessary for the successful, ethical adoption of the powerful technological solutions.
The Personalized Multimodal Treatment Response System is a unique framework due to the described Continuous Learning and Adaptation module, which introduced a unique feedback loop. It is used to continually update the developed phenomenological model of a patient based on new data and outcomes. Therefore, such a design makes the system a perfect representation of a continuously evolving one. Furthermore, the presented framework highlights the issues of patients' data loss or unauthorized access by implementing strict integrated data security. Therefore, the combination of continuous learning and adaptation and restrictions on data access or improvements is a clear example of balancing the advantages of personalized medicine and protecting sensitive information.. However, one of the low side is associated with an inability to take advantage of some data sources to make treatment predictions more precise, thereby reducing the overall model accuracy. This is primarily caused by the described heterogeneity and complexity of integrated data types, where structured genetic data are combined with unstructured medical images or electronic health records. To do that, the accelerated loss of features or over-simplification caused by either early fusion, late fusion, or predefined model-based integration is possible.. In terms of the second limitation presented above, it is also worth mentioning different potential biases of the data that are provided to the model, making it difficult to predict the relevance of the suggestions10. In such a way, it is necessary to conclude that the presented framework can be perceived as a reliable and efficient solution, even though it has two drawbacks. They are connected to the scale and representativeness of the used data and the ability to balance the need for model adaptation and privacy protection.

References

[1] M. Jaiswal and E. Mower Provost, "Privacy enhanced multimodal neural representations for emotion recognition," *Proc. Conf. AAAI Artif. Intell.*, vol. 34, no. 05, pp. 7985–7993, Apr. 2020.

[2] P. Chen, Q. Li, D. Zhang, Y. Yang, and Z. Cai, "A survey of multimodal machine learning," *Chinese Journal of*, 2020.

[3] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova, "Supervised machine learning in multimodal learning analytics for estimating success in project-based learning," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 366–377, Aug. 2018.

[4] Y. Ren, N. Xu, M. Ling, and X. Geng, "Label distribution for multimodal machine learning," *Frontiers of Computer Science*, vol. 16, no. 1, p. 161306, Sep. 2021.

[5] T. Mitchell, "Learning Analytics as AI Extenders in Education: Multimodal Machine Learning versus Multimodal Learning Analytics," in *Proceedings of AIAED 2019*, vol. 2019, T. Mitchell, Ed. Beijing, China: AIAED, 2019.

[6] N. Koutsouleris *et al.*, "Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression," *JAMA Psychiatry*, vol. 78, no. 2, pp. 195–209, Feb. 2021.

[7] A. Barua, M. U. Ahmed, and S. Begum, "A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions," *IEEE Access*, vol. 11, pp. 14804–14831, 2023.

[8] L. Boussioux, C. Zeng, T. Guénais, and D. Bertsimas, "Hurricane Forecasting: A Novel Multimodal Machine Learning Framework," *Weather Forecast.*, vol. 37, no. 6, pp. 817–831, Jun. 2022.

[9] A. K. Saxena, R. R. Dixit, and A. Aman-Ullah, "An LSTM Neural Network Approach to Resource Allocation in Hospital Management Systems," *International Journal of Applied*, 2022.

[10] M. Moshawrab, M. Adda, A. Bouzouane, H. Ibrahim, and A. Raad, "Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection," *Electronics*, vol. 12, no. 7, p. 1558, Mar. 2023.

[11] A. K. Saxena, "Balancing Privacy, Personalization, and Human Rights in the Digital Age," *Eigenpub Review of Science and Technology*, vol. 4, no. 1, pp. 24–37, 2020.

[12] Y. Xia, Z. Ji, A. Krylov, H. Chang, and W. Cai, "Machine Learning in Multimodal Medical Imaging," *Biomed Res. Int.*, vol. 2017, p. 1278329, Mar. 2017.

[13] R. Walambe, P. Nayak, A. Bhardwaj, and K. Kotecha, "Employing Multimodal Machine Learning for Stress Detection," *J. Healthc. Eng.*, vol. 2021, p. 9356452, Oct. 2021.

[14] K. Xu *et al.*, "Multimodal Machine Learning for Automated ICD Coding," in *Proceedings of the 4th Machine Learning for Healthcare Conference*, 2019, vol. 106, pp. 197–215.

[15] A. Tiulpin *et al.*, "Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data," *Sci. Rep.*, vol. 9, no. 1, p. 20038, Dec. 2019.

[16] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions," *arXiv [cs.LG]*, 07-Sep-2022.