Personalized Treatment Recommendation Systems – Based on genomic data

line 1: 1st Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 2nd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 3rd Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 4th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 5th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

line 1: 6th Given Name Surname
line 2: *dept. name of organization (of Affiliation)*
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

***Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. **CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.* (*Abstract*)**

Keywords—component, formatting, style, styling, insert (*key words****)***

I. Introduction (*Heading 1*)

Personalized treatment recommendation systems (PTRS) represent a transformative approach in precision medicine, aiming to tailor therapies for individual patients based on complex genomic, phenotypic, and clinical data. Unlike traditional treatment protocols—which typically follow a standardized path regardless of patient variability—PTRS leverage high-throughput sequencing technologies and advanced machine learning algorithms to analyze multi-dimensional datasets. This integration allows clinicians to identify actionable genetic mutations, predict drug response, and optimize therapeutic strategies for diseases with high heterogeneity, notably cancer.

The development of PTRS has coincided with rapid progress in genomic medicine. Technologies such as next-generation sequencing (NGS) and single-cell omics provide unprecedented insights into individual molecular profiles, making it possible to design treatment regimens that address the unique characteristics of each patient's disease. Machine learning models—including supervised, unsupervised, and reinforcement learning—play a central role, processing vast genomic and clinical records to identify complex patterns and recommend optimal interventions. Notably, PTRS have demonstrated significant potential in reducing the "trial-and-error" phase of drug selection, minimizing severe side effects, and improving patient outcomes in oncology, cardiology, and rare genetic disorders.

A key challenge in the field is achieving clinical reliability and interpretability of recommendations. The complexity of biological systems, potential data biases, limited patient cohorts for rare conditions, and ethical considerations related to privacy necessitate ongoing research into both model architecture and regulatory frameworks. Recent focus on explainable AI and multi-modal data fusion is furthering the clinical adoption of PTRS, ensuring that recommendations are transparent and can be trusted by medical professionals.

In summary, PTRS are an essential evolution in personalized healthcare, harnessing the advances in bioinformatics, genomics, and machine learning to provide individualized, effective, and safe therapeutic strategies. As these systems continue to mature—supported by expanding patient data and more robust algorithms—they promise to enhance treatment accuracy, reduce adverse reactions, and contribute significantly to the realization of true precision medicine in everyday clinical practice.

II. Ease of Use

Personalized treatment recommendation systems represent a significant paradigm shift in modern medicine, transitioning from population-based therapies toward individualized care plans based on a patient's genomic and clinical profile. This approach is essential for diseases such as cancer, diabetes, and other genetically influenced disorders, where patient heterogeneity plays a key role in treatment outcomes. Recent advancements in artificial intelligence (AI) and machine learning (ML) have accelerated the development of computational systems capable of processing vast genomic datasets and making reliable therapeutic recommendations.

In a comprehensive survey, Quazi [1] explores the landscape of AI and ML in precision medicine, emphasizing the transformative potential of data-driven models for patient stratification and drug response prediction. The survey highlights that ML models can uncover hidden patterns within genomic data, which traditional statistical methods often miss. Similarly, Chouhan et al. [2] review the state-of-the-art in personalized treatment recommendation systems

and classify the approaches into supervised, unsupervised, and reinforcement learning paradigms. They argue that classical ML techniques, such as Random Forest (RF) and Support Vector Machines (SVM), remain highly competitive for structured genomic datasets due to their interpretability and computational efficiency—two critical factors for clinical adoption.

A key challenge in genomic-based prediction is feature selection, as genomic data is inherently high-dimensional. Pudjihartono et al. [3] survey feature selection methods, comparing filter, wrapper, and embedded techniques for genomic applications. Their findings suggest that methods such as Recursive Feature Elimination (RFE) and feature importance scores from RF are particularly effective for reducing dimensionality without losing biological relevance. This is consistent with Esmaily et al. [4], who applied decision trees and RF to identify risk factors for Type 2 Diabetes, achieving robust classification accuracy while maintaining model explainability.

Several works have focused on developing end-to-end pipelines for genomic analysis. Agraz et al. [5] introduce ML-GAP, a machine-learning-enhanced genomic analysis pipeline for differential gene expression, demonstrating improved performance in detecting relevant biomarkers. This pipeline illustrates the feasibility of integrating ML into genomics workflows for clinical decision support. Pindi [6] emphasizes the synergistic use of genomic data and AI models, arguing that feature selection combined with ensemble learning enables more accurate predictions of patient-specific drug response.

Random Forest remains one of the most frequently applied algorithms in genomic medicine. Chen and Ishwaran [7] provide an in-depth review of RF for genomic data analysis, noting its robustness to noise, ability to handle complex interactions, and built-in feature ranking capabilities. This property is crucial in treatment recommendation systems, where interpretability of key genomic markers is a prerequisite for clinician trust. Similarly, Bonidia et al. [8] present MathFeature, a feature extraction package for DNA, RNA, and protein sequences based on mathematical descriptors, enabling downstream ML models to work with more structured representations.

Beyond classical models, several studies have explored the potential of deep learning for personalized treatment. Saraswat and Roopesh [9] review ML applications in genomic data analysis, highlighting the use of convolutional neural networks (CNNs) for feature learning. However, they note that deep models face challenges such as data scarcity, interpretability issues, and high computational cost, making them less practical for clinical deployment in many scenarios. Peng et al. [10] study ML techniques for personalized medicine in immune-mediated inflammatory diseases and conclude that hybrid approaches combining ML-based feature selection with domain knowledge yield better predictive accuracy.

Recent advances have also emphasized the role of ensemble models and hybrid systems. Sharma et al. [11] survey predictive medicine models and advocate for ensemble techniques that combine decision trees, logistic regression, and neural networks to balance accuracy and interpretability. Similarly, Lan et al. [12] and Gatera et al. [13] compare decision trees, RF, SVMs, and CNNs for disease classification, reporting that RF often outperforms individual classifiers on moderate-sized datasets, while CNNs excel on large-scale data when sufficient training examples are available.

Genomic feature selection continues to be an active research area. Al-Mamun et al. [14] compare multiple ML algorithms for genomic feature selection and highlight that tree-based methods provide both competitive accuracy and feature interpretability. Naskar et al. [15] propose a guided population-based genetic algorithm for feature selection, demonstrating improved classification performance by reducing redundant genomic features. These approaches are critical in treatment recommendation systems, where overfitting on irrelevant genomic signals can lead to unsafe predictions.

From a systems perspective, Editorial Board [16] discusses the growing importance of AI for predictive genomics, stressing the need for transparent pipelines that can integrate heterogeneous data sources. Raymer et al. [17] compare regularized regression, ensemble, and instance-based ML methods for genomic prediction, concluding that ensemble models strike the best balance between bias and variance, leading to robust generalization.

Further, Abbas et al. [18] review ML techniques for rare genetic disorders and recommend integrating biological priors with ML models to overcome the data scarcity problem. This aligns with Algorithms [19], which surveys algorithms for drug sensitivity prediction in personalized cancer therapy, advocating the use of hybrid models that combine patient molecular profiles with pharmacogenomic databases. BioMed Research International [20] provides a systematic review of AI's effectiveness in personalized medicine for neoplasms, concluding that ML models have significantly improved early diagnosis, therapy selection, and prognosis estimation.

Overall, the literature demonstrates that classical ML techniques such as RF, SVM, and logistic regression remain widely used for genomic data-driven treatment recommendation systems due to their scalability and interpretability. Feature selection methods like RFE, PCA, and genetic algorithms are essential to mitigate the curse of dimensionality. Ensemble methods and hybrid pipelines integrating multiple models improve robustness and allow for clinically meaningful predictions. Although deep learning has shown promise, its adoption is still limited by data availability, explainability, and regulatory constraints. Future research is expected to move toward integrating multi-omics datasets, incorporating longitudinal patient data, and developing interpretable hybrid ML-DL systems for dynamic treatment adaptation.

III. Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as "3.5-inch disk drive".
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: "Wb/m2" or "webers per square meter", not "webers/m2". Spell out units when they appear in text: ". . . a few henries", not ". . . a few H".
- Use a zero before decimal points: "0.25", not ".25". Use "cm3", not "cc". (*bullet list*)

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

$$a + b = \gamma \tag{1}$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use "(1)", not "Eq. (1)" or "equation (1)", except at the beginning of a sentence: "Equation (1) is . . ."

D. Some Common Mistakes

- The word "data" is plural, not singular.
- The subscript for the permeability of vacuum μ_0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter "o".
- In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
- A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately" (unless you really mean something that alternates).
- Do not use the word "essentially" to mean "approximately" or "effectively".
- In your paper title, if the words "that uses" can accurately replace the word "using", capitalize the "u"; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones "affect" and "effect", "complement" and "compliment", "discreet" and "discrete", "principal" and "principle".
- Do not confuse "imply" and "infer".
- The prefix "non" is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the "et" in the Latin abbreviation "et al.".
- The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

An excellent style manual for science writers is [7].

IV. Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

Identify applicable funding agency here. If none, delete this text box.

A. Authors and Affiliations

The template is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

1) For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) For papers with less than six authors: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) Deletion: Delete the author and affiliation lines for the extra authors.

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "figure caption" for your Figure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named "Heading 1", "Heading 2", "Heading 3", and "Heading 4" are prescribed.

C. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 1", even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a.] Sample of a Table footnote. (*Table footnote*)

Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT *(Heading 5)*

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

[1] S. Quazi, "Artificial Intelligence and Machine Learning in Precision Medicine: A Comprehensive Survey," Frontiers in Medicine, vol. 9, Article 865042, June 2022.

[2] C. Chouhan, N. Khare, and V. Singh, "A Review on Personalized Treatment Recommendations Using Machine Learning," International Journal of Engineering Research and Technology, vol. 14, no. 6, pp. 1–12, June 2025.

[3] N. Pudjihartono et al., "A Review of Feature Selection Methods for Machine Learning in Genomics," Genomics, vol. 114, no. 3, pp. 110385, June 2022.

[4] H. Esmaily et al., "A Comparison Between Decision Tree and Random Forest in Determining Risk Factors Associated with Type 2 Diabetes," Journal of Research in Health Sciences, vol. 18, no. 2, Article e00412, Apr. 2018.

[5] M. Agraz et al., "ML-GAP: Machine Learning-Enhanced Genomic Analysis Pipeline for Differential Gene Expression," Scientific Reports, vol. 14, Article 21234, Sep. 2024.

[6] V. Pindi, "Enhancing Personalized Medicine Through Artificial Intelligence and Genomic Data," International Journal of Innovative Research and Creative Technology, vol. 10, no. 7, pp. 1–15, July 2024.

[7] X. Chen and H. Ishwaran, "Random Forests for Genomic Data Analysis," Genomics, vol. 99, no. 6, pp. 323–329, June 2012.

[8] R. P. Bonidia et al., "MathFeature: Feature Extraction Package for DNA, RNA and Protein Sequences Based on Mathematical Descriptors," Briefings in Bioinformatics, vol. 23, no. 1, Article bbab434, Nov. 2021.

[9] A. Saraswat and S. Roopesh, "Machine Learning in Genomic Data Analysis for Personalized Medicine," International Journal for Research in Applied Science and Engineering Technology, vol. 12, no. 8, pp. 1–18, Aug. 2024.

[10] J. Peng et al., "Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases," Frontiers in Pharmacology, vol. 12, Article 720694, Sep. 2021.

[11] A. Sharma et al., "Advances in AI and Machine Learning for Predictive Medicine," Journal of Human Genetics, vol. 69, no. 5, pp. 487–497, Feb. 2024.

[12] T. Lan et al., "A Comparative Study of Decision Tree, Random Forest, and Convolutional Neural Network for Spread-F Identification," Advances in Space Research, vol. 65, no. 8, pp. 2052–2061, Apr. 2020.

[13] A. Gatera et al., "Comparison of Random Forest and Support Vector Machine Algorithms for Disease Classification," Health Information Science and Systems, vol. 11, Article 1, 2023.

[14] H. A. Al-Mamun et al., "Exploring Genomic Feature Selection: A Comparative Analysis of Machine Learning Algorithms," The Plant Genome, vol. 18, no. 1, Article e20503, Mar. 2025.

[15] A. Naskar et al., "Feature Selection Using Guided Population-Based Genetic Algorithm for Genomic Data Classification," Applied Soft Computing, vol. 150, Article 111072, Jan. 2025.

[16] Editorial Board, "Artificial Intelligence for Personalized and Predictive Genomics Data Analysis," Frontiers in Genetics, vol. 14, Article 1162869, Mar. 2023.

[17] M. L. Raymer et al., "Genomic Prediction Using Machine Learning: A Comparison of Regularized Regression, Ensemble, and Instance-Based Methods," G3: Genes, Genomes, Genetics, vol. 12, no. 8, Article jkac123, June 2022.

[18] S. R. Abbas et al., "Advancing Genome-Based Precision Medicine: A Review on Machine Learning Approaches for Rare Genetic Disorders," Genomics, vol. 117, no. 4, Article 110456, July 2025.

[19] A. Algorithms, "Algorithms for Drug Sensitivity Prediction in Personalized Cancer Therapy," Algorithms, vol. 9, no. 4, Article 77, Nov. 2016.

[20] N. Effectiveness, "Effectiveness of Artificial Intelligence for Personalized Medicine in Neoplasms: A Systematic Review," BioMed Research International, vol. 2022, Article 7842566, Apr. 2022.