Journal of Computational
Intelligence and Decision Analytics

Journal homepage: www.cida-journal.org
ISSN: 3102-6125

Computational Prioritization of Bioactive Molecules for Therapeutic and Nutritional Design Using ML–MCDM Approaches

Salma Kanwal,[1] Khadija Sarwar[1,*]

[1] Department of Mathematics, Lahore College for Women University, Lahore, Pakistan

ARTICLE INFO

Article history:
Received 5 June 2026
Received in revised form 29 July 2026
Accepted 24 August 2026
Available online 26 August 2026

Keywords:
Topological indices; Acyclic structures; MCDM; Machine Learning; PN solutions

ABSTRACT

Parenteral Nutrition (PN) solutions are specialized intravenous formulations that provide vital nutrients (amino acids) to patients who are unable to meet their nutritional needs orally or enterally. This study emphasizes the significance of acyclic amino acids, which are major components of Parenteral Nutrition (PN) solutions, as they are linear in structure and have effective metabolic applications in clinical nutrition. The present study combines machine learning and multi-criteria decision-making techniques (MCDM), and this combination of approaches is used to rank acyclic amino acids using entropy-based topological indices. This framework is based on regression techniques such as linear, polynomial, LASSO, and Random Forest regression and uses entropy measures of Gourava indices as features. To rank acyclic amino acids, machine learning tools are employed in parallel with MCDM tools (VIKOR and EDAS). The outcomes linked to rankings are tabulated and visualized by means of bar plots and heatmaps to ensure interpretability and transparency. Graphical comparisons and correlation analysis were used to confirm that the ML and MCDM ranks were consistent. Furthermore, the computational rankings are compared to three commercially available PN solutions (Aminoven, TrophAmine, and ProSol) to ensure their clinical relevance. This integration enabled us to compare amino acid priority using ML-MCDM techniques to real-world formulations utilized in parenteral nutrition.

1. Introduction

Intravenous formulations known as Parenteral Nutrition (PN) solutions are intended to meet the full nutritional requirements of patients who are unable to be fed enterally or orally. Amino acids, the main supply of nitrogen and necessary for tissue repair, protein synthesis, and general metabolic processes, are among the main ingredients of PN solutions. This study focuses on a specific type of amino acid: those having acyclic structures. An amino acid with a tree-like molecular structure is said to be acyclic if its carbon skeleton is free of aromatic or cyclic rings. These amino acids have more flexible geometries and lack resonance stability from rings; therefore, when assessing indices that are more sensitive to branching, path lengths, and atomic connectivity, these acyclic structures make graph-theoretic analysis easier to understand. The efficient analysis and ranking of amino acids according to their structural and physicochemical properties can assist in prioritizing amino acids for

*Corresponding author.
E-mail address: Khadijasarwar.khadijasarwar24@gmail.com

specific applications and enhance their functional significance.

Topological indices that are graph-based descriptors are employed to numerically represent these molecular structures.

Entropy-based topological indices have drawn interest in addition to classical indices because of their capacity to quantify the structural complexity and uncertainty of molecular graphs. Information theory inspired entropy metrics, like Shannon entropy by Shannon [1], offer an alternative viewpoint by expressing the distribution of atoms and bonds in terms of probability dispersion. These metrics may be determined by distance, degree, or other topological indices. In the late 1950s, Rashevsky [2] established the foundation for graph entropy research by becoming the first to utilize entropy to quantify the structural information content of graphs.

Numerous disciplines, including network analysis, chemistry, and drug development, use topological indices and the entropies that go along with them. They support the analysis of complex systems, the prediction of molecular properties, and the comprehension of structure-activity correlations.

Graph entropy measurements, whether intrinsic or extrinsic, are used to quantify structural complexity. Probability distributions are linked to network elements such as edges and vertices through intrinsic measures [3]. As information functionals, degree powers and other important invariants are frequently employed. These functionals are useful for capturing structural insights, as demonstrated by Dehmer's work on graph entropies based on them see [4-6].

Assume $G = (V(G), E(G), W(xy))$ is a connected graph, with $\rho =$ the order and $\varrho =$ the size of the graph. Then $V(G)$ and $E(G)$ signify its vertex and edge sets, respectively, and $W(xy)$ signifies a meaningful function. The number of edges connected to a vertex x is the degree of that vertex, and $h(x)$ is the degree of x from a closed neighborhood. If two vertices x and y are connected by an edge, say $e = xy$, they are neighbors.

Now we can define the entropy function as follows for a graph G. The entropy measures of Gourava indices are presented in [7].

$$ENT_W(G) = -\sum_{i=1}^{\rho} \frac{W(x_i)}{\sum_{i=1}^{\rho} W(x_i)} \log \left[\frac{W(x_i)}{\sum_{i=1}^{\rho} W(x_i)} \right] \tag{1}$$

Now if $W(x_i) = h(x_i)$ as $W(x_i)$ denotes the degree of x_i and $x_i \in V$, then equation [1] becomes

$$ENT_W(G) = -\sum_{i=1}^{\rho} \frac{h(x_i)}{\sum_{j=1}^{\rho} h(x_j)} \log \left[\frac{h(x_i)}{\sum_{j=1}^{\rho} h(x_j)} \right]$$

$$ENT_W(G) = \log \left(\sum_{j=1}^{\rho} h(x_j) \right) - \frac{1}{\sum_{j=1}^{\rho} h(x_j)} \sum_{i=1}^{\rho} [h(x_i) \log(h(x_i))]$$

By fundamental theorem of graph theory $\sum_{j=1}^{\rho} h(x_j) = 2y$, thus we have

$$ENT_W(G) = \log (2y) - \frac{1}{2y} \log \left[\sum_{i=1}^{\rho} h(x_i)^{h(x_i)} \right] \tag{2}$$

Chen *et al.* [8] presented edge-weighted graph entropy, for some graph $G = (V(G); E(G); W(xy))$. For this, the entropy function is defined as

$$ENT_W(G) = -\sum_{x'y' \in E(G)} \frac{W(x'y')}{\sum_{xy \in E(G)} W(xy)} \log \left[\frac{W(x'y')}{\sum_{xy \in E(G)} W(xy)} \right] \tag{3}$$

First Gourava entropy:

If $W(xy) = h(x) + h(y) + h(x)h(y)$, then

$$\sum_{xy \in E(G)} W(xy) = \sum_{xy \in E(G)} (h(x) + h(y) + h(x)h(y)) = GO_1(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{GO_1}(G) = \log\left(GO_1(G)\right) - \frac{1}{(GO_1(G))}\log\left[\sum_{xy\in E(G)}[h(x)+h(y)+h(x)h(y)]^{[h(x)+h(y)+h(x)h(y)]}\right] \quad (4)$$

Second Gourava entropy:

If $W(xy) = (h(x)^2h(y) + h(x)h(y)^2$, then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} h(x)^2h(y) + h(x)h(y)^2 = GO_2(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{GO_2}(G) = \log\left(GO_2(G)\right) - \frac{1}{(GO_2(G))}\log\left[\sum_{xy\in E(G)}\left[h(x)^2h(y)+h(x)h(y)^2\right]^{\left[h(x)^2h(y)+h(x)h(y)^2\right]}\right] \quad (5)$$

First Hyper Gourava entropy:

If $W(xy) = (h(x) + h(y) + h(x)h(y))^2$, then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} (h(x)+h(y)+h(x)h(y))^2 = HGO_1(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{HGO_1}(G) = \log\left(HGO_1(G)\right) - \frac{1}{(HGO_1(G))}\log\left[\sum_{xy\in E(G)}\left[(h(x)+h(y)+h(x)h(y))^2\right]^{\left[(h(x)+h(y)+h(x)h(y))^2\right]}\right] \quad (6)$$

Second Hyper Gourava entropy:

If $W(xy) = (h(x)^2h(y) + h(x)h(y)^2)^2$ then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} (h(x)^2h(y)+h(x)h(y)^2)^2 = HGO_2(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{HGO_2}(G) = \log\left(HGO_2(G)\right) - \frac{1}{(HGO_2(G))}\log\left[\sum_{xy\in E(G)}(h(x)^2h(y)+h(x)h(y)^2)^2\right]^{\left[(h(x)^2h(y)+h(x)h(y)^2)^2\right]} \quad (7)$$

First Alpha Gourava entropy:

If $W(xy) = (h(x)^2 + h(y)^2 + h(x)h(y))$ then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} h(x)^2 + h(y)^2 + h(x)h(y) = AGO_1(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{AGO_1}(G) = \log\left(AGO_1(G)\right) - \frac{1}{(AGO_1(G))}\log\left[\sum_{xy\in E(G)}\left[h(x)^2+h(y)^2+h(x)h(y)\right]^{\left[h(x)^2+h(y)^2+h(x)h(y)\right]}\right] \quad (8)$$

Second Alpha Gourava entropy:

If $W(xy) = h(x)^3h(y) + h(x)h(y)^3$, then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} h(x)^3h(y) + h(x)h(y)^3 = AGO_2(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{AGO_2}(G) = \log(AGO_2(G)) - \frac{1}{AGO_2(G)} \log\left[\sum_{xy\in E(G)} \left[h(x)^3h(y) + h(x)h(y)^3\right]^{\left[h(x)^3h(y)+h(x)h(y)^3\right]}\right] \quad (9)$$

First Gamma Gourava entropy:
If $W(xy) = h(x)^3 + h(y)^3 + h(x)h(y)$, then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} h(x)^3 + h(y)^3 + h(x)h(y) = GA(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{GA}(G) = \log(ABC(G)) - \frac{1}{GA(G)} \log\left[\sum_{xy\in E(G)} \left[h(x)^3 + h(y)^3 + h(x)h(y)\right]^{\left[h(x)^3+h(y)^3+h(x)h(y)\right]}\right] \quad (10)$$

Second Gamma Gourava entropy:
If $W(xy) = h(x)^4h(y) + h(x)h(y)^4$, then

$$\sum_{xy\in E(G)} W(xy) = \sum_{xy\in E(G)} h(x)^4h(y) + h(x)h(y)^4 = GGO_2(G)$$

By using Equation 1, we get the formula for first Zagreb entropy as follows:

$$ENT_{GGO_2}(G) = \log\left(GGO_2(G)\right) - \frac{1}{(GGO_2(G))} \log\left[\sum_{xy\in E(G)} \left[h(x)^4h(y) + h(x)h(y)^4\right]^{\left[h(x)^4h(y)+h(x)h(y)^4\right]}\right] \quad (11)$$

A systematic and quantitative approach to evaluating many conflicting criteria is provided by Multi-Criteria Decision Making (MCDM) techniques. Many efforts have been made to use MCDM approaches to rank or study the medications [9]. Taherdoost [10] offers a comprehensive lesson for ranking and decision-making in multi-attribute contexts, along with an explanation of the SAW technique. Using QSPR analysis and topological indices, the study in [11] used decision-making techniques to establish the manufacturing sequence of several medications used to treat bone cancer.

Pharmaceutical compounds have been evaluated and ranked using multicriteria decision-making (MCDM) methodologies in conjunction with QSAR and QSPR models. In [12], QSAR, QSPR, and structural characteristics were integrated to promote optimal drug selection while ranking asthma drugs using VIKOR. Similarly, using TOPSIS and VIKOR for ranking, [13] employed multiple linear regression to connect physicochemical properties with topological indices for dry eye treatments. TOPSIS and SAW were used in the study in [14] to rank eye drugs and encourage well-informed decisions. In [15], medicines for lung diseases were analyzed using degree-based topological indices to correlate their physical properties with their structures. These studies show how integrating structural analysis, decision-making techniques, and predictive modeling facilitates the assessment and creation of new drugs.

Husin *et al.* [16] examined kidney cancer medicines using topological indices. With the QSPR model being the most appropriate for forecasting properties and achieving the best fit, the study presented in [16] shows a strong correlation between TIs and physical qualities. Additionally, they used multicriteria decision-making tools such as COPRAS and PROMETHEE-II to rank the physicochemical properties and medications used to treat the medical condition.

At the same time, machine learning (ML) techniques provide strong tools for feature selection and pattern detection. Topological descriptors and entropy of these drugs were calculated for analyzing

the structural properties, in order to enhance our understanding of molecular behavior. Furthermore, their physicochemical properties were explored by utilizing supervised machine learning algorithms and performing Quantitative Structure-Property Relationship (QSPR) analysis, which explains the connections between the topological descriptor and physicochemical attributes [17].

The study in [18] emphasizes how MCDM techniques can be combined with machine learning (ML) algorithms to enhance decision-making. Thyroid illness is analyzed employing Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR). Choosing the most pertinent topological indices becomes crucial as the number of them rises. MCDM reduces dimensionality, improves interpretability, and helps find the most important indices. Hybrid MCDM–ML techniques have also been investigated in recent research. For instance, utilizing machine learning and eigenvalue based indices, Shanmukha *et al.* [19] created a QSPR model for benzenoid hydrocarbons. The SAW approach was utilized to rank the top molecules.

This study in [20] offers a thorough investigation of the quantitative structure-property relationship (QSPR) pertaining to cancer medications, emphasizing the use of data analytic techniques and topological indices (TI). In addition to linear regression model, two other machine learning models, SVR and Random Forest, were also used for additional analysis and comparison of their performance in predicting the physicochemical properties of drugs, in order to evaluate the benefits and drawbacks of each model. The linear regression model was the model that performed the best.

In this work we employed Machine Learning regression models including linear and polynomial regression, Random Forest (RF) and Lasso regression to predict key physicochemical properties (e.g., molecular weight, melting point, isoelectric point, and molecular volume) from topological indices.

By combining the interpretability of MCDM with the predicting features of machine learning, this study designs a combination of approaches to rank acyclic amino acids while offering the significant insights into molecular design, property prediction, and the integration of structure-based decision tools.

For drug design the suggested ML–MCDM approach provides a fresh and methodical way to rank amino acids according to their structural and physicochemical characteristics. This work is applicable in the formulation of parenteral nutrition (PN) solutions based on amino acids, where balanced molecular profiles are essential for patient care and in medication design, where the best choice of amino acids affects molecular efficacy. This study allows researchers to prioritize amino acids for particular biochemical activities by finding the most important topological characteristics and combining predictive modeling with decision-making procedures.

2. Methodology

This section contains all the data used in this study and provides the methodology to perform calculations.

2.1 Dataset Description

The dataset consists of 13 acyclic structures of amino acids, each characterized by a set of topological indices and physicochemical properties. The topological indices used in this study include classical and entropy-based graph invariants derived from molecular graph representations of the considered acyclic structures.

All indices and properties were extracted from published literature or calculated using molecular graph descriptors.

The study excludes any cyclic or aromatic substructures and concentrates on thirteen acyclic amino acid compounds that were chosen for their linear chain properties. Glycine, Alanine, Valine, Asparagine, Leucine, Serine, Threonine, Aspartic acid, Glutamic acid, Glutamine, Lysine, Methionine and Arginine are some of these structures. In this study 2D structures are taken from *PubChem*.

The entropy measure values for the amino acids are represented by the matrix entries presented in Table 1. The considered physicochemical properties of acyclic amino acids are presented in Table 2.

The following target properties are taken into account for ranking and predictive modeling.

Table 1
Entropy Measures of Topological indices computed

Amino Acids	ENT_{GO_1}	ENT_{GO_2}	ENT_{HGO_1}	ENT_{HGO_2}	ENT_{AGO_1}	ENT_{AGO_2}	ENT_{GGO_1}	ENT_{GGO_2}
Glycine	2.094676	1.826040	1.806042	0.854671	2.078736	1.71380	2.02660	1.639451
Alanine	2.31063	2.07055	2.038125	1.221808	2.360922	1.954903	2.319523	1.882517
Serine	2.576774	2.144764	2.115339	1.342284	2.419674	2.019694	2.360816	1.938206
Aspartic	2.583696	2.288261	2.259824	1.527970	2.563145	2.157344	2.501443	2.072854
Asparagine	2.655931	2.52149	2.345933	1.631958	2.621316	2.214057	2.59886	2.167523
Threonine	2.645052	2.346385	2.313102	1.607592	2.633486	2.231613	2.587740	2.162297
Glutamic Acid	2.76335	2.46515	2.433070	1.749130	2.750322	2.343886	2.699557	2.264290
Valine	2.767168	2.469106	2.43274	1.758739	2.772799	2.370506	2.749544	2.316897
Methionine	2.842969	2.580591	2.546589	1.800319	2.846860	2.48745	2.824880	2.435207
Glutamine	2.824087	2.651706	2.504637	1.828129	2.799381	2.387430	2.778881	2.344678
Leucine	2.920764	2.625365	2.588661	1.951159	2.930338	2.521495	2.912821	2.485116
Lysine	3.008867	2.714362	2.680096	2.053061	3.146799	2.616937	3.115198	2.562617
Arginine	3.099495	2.813243	2.784102	2.189654	3.086667	2.699200	3.338542	2.616717

Table 2
Physicochemical Properties

Amino Acids	MW (Da)	MP (°C)	PI (pI)	pKa_1 ($\alpha - COOH$)	pKa_2 ($\alpha - NH_2$)	$Volume$ (A^3)
Glycine	75.07	233	5.97	2.34	9.6	60.1
Alanine	89.09	258	6.01	2.34	9.69	88.6
Serine	105.09	246	5.68	2.21	9.15	89
Aspartic Acid	133.1	270	2.77	1.88	9.6	111.1
Asparagine	132.12	234	5.41	2.1	8.8	114.1
Threonine	119.12	256	5.6	2.09	9.1	116.1
Glutamic Acid	147.13	199	3.22	2.19	9.67	138.4
Valine	117.15	298	5.96	2.32	9.62	140
Methionine	149.21	281	5.74	2.28	9.21	162.9
Glutamine	146.15	185	5.65	2.17	9.13	143.8
Leucine	131.18	293	5.98	2.36	9.6	166.7
Lysine	146.19	224.5	9.74	2.18	8.95	168.6
Arginine	174.2	260	10.76	2.17	9.04	173.4

2.2 Methodology

In order to rank acyclic amino acids according to anticipated physicochemical attributes and structural indices, this study uses an integrated Machine Learning (ML) and Multi-Criteria Decision Making (MCDM) framework. Each of the dataset's 13 acyclic amino acids is identified by a set of eight topological indices, which include entropy-based metrics like $Ent_{GO_1}, Ent_{GO_2}, Ent_{HGO_1}, Ent_{HGO_2}, Ent_{AGO_1}$, Ent_{AGO_2}, Ent_{GGO_1} and Ent_{AGO_2}. In this study, we performed traditional regression models (linear and polynomial) and machine learning regression models, i.e., Random Forest (RF) and Lasso regression models. Entropy measures of Gourava indices are used as input features. Molecular weight, molecular volume, melting point,pka_1, pka_2 and isoelectric point are the target variables for prediction. Each property is modeled distinctly to analyze how well these structural indices could explain property behavior. For each regression model, predictions for all six physicochemical properties

were derived and normalized by employing min-max scaling for optimally comparable results. As the dataset is small, exploratory data analysis (EDA) is restricted to basic statistical checks and correlation analysis. For each model and property, we produced significant regression metrics such as the R^2 score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to better analyze model performance. These regression outcomes ensure that topological indices encode significant structural characteristics; hence, they can be employed to predict quantitative structure-property relationships (QSPR) in amino acids. The research compiled data on the concentrations of acyclic amino acids from different commercially available PN solutions (Aminoven, TrophAmine, and ProSol). Based on these data, the relative significance of each acyclic amino acid among PN concentrations is obtained by comparing them with the MCDM-ML ranks. All calculations and visuals are derived using Python libraries including scikit-learn, pandas, and matplotlib in a Google Colab environment.

3. Results and Discussions

3.1 Exploratory Data Analysis (EDA)

A crucial initial phase of every data-driven study is exploratory data analysis (EDA), which is meant to understand the dataset's basic framework, trends, and relations. It refers to merging both statistical and visual tools to summarize the main features of the data. The distribution, central tendency, and variability of the topological indices and physicochemical characteristics of amino acids are investigated in this work using EDA. Strong linear relationships between entropy measures and attributes like molecular weight and molecular volume are found using correlation analysis, while summary statistics like mean, standard deviation, and range shed light on the magnitude and distribution of each variable, as shown in Table 3 and Figure 1. Additionally, strongly associated features and any redundancies are visually identified using a correlation heatmap. This early study guided the selection of machine learning models for property prediction and supported the use of topological indices (entropy measures) as predictive factors. All things considered, EDA made sure the data was appropriate for modeling and allowed for an insightful understanding of the connections between physicochemical behavior and amino acid structure.

Table 3
Summary Statistics of EDA

Feature/Property	Mean	Std Dev	Min	Max
$Ent(GO_1)$	2.7	0.27	2.09	3.1
$Ent(GO_2)$	2.42	0.28	1.83	2.81
$Ent(HGO_1)$	2.37	0.27	1.81	2.78
$Ent(HGO_2)$	1.66	0.36	0.85	2.19
$Ent(AGO_1)$	2.69	0.3	2.08	3.15
$Ent(AGO_2)$	2.29	0.28	1.71	2.7
$Ent(GGO_1)$	2.68	0.34	2.03	3.34
$Ent(GGO_2)$	2.22	0.29	1.64	2.62
MW	128.06	26.93	75.07	174.2
MP	249.04	33.84	185	298
PI	6.04	2.15	2.77	10.76
$pKa_1(\alpha - COOH)$	2.2	0.13	1.88	2.36
$pKa_2(\alpha - NH_2)$	9.32	0.32	8.8	9.69
$Volume(A^3)$	128.68	35.62	60.1	173.4

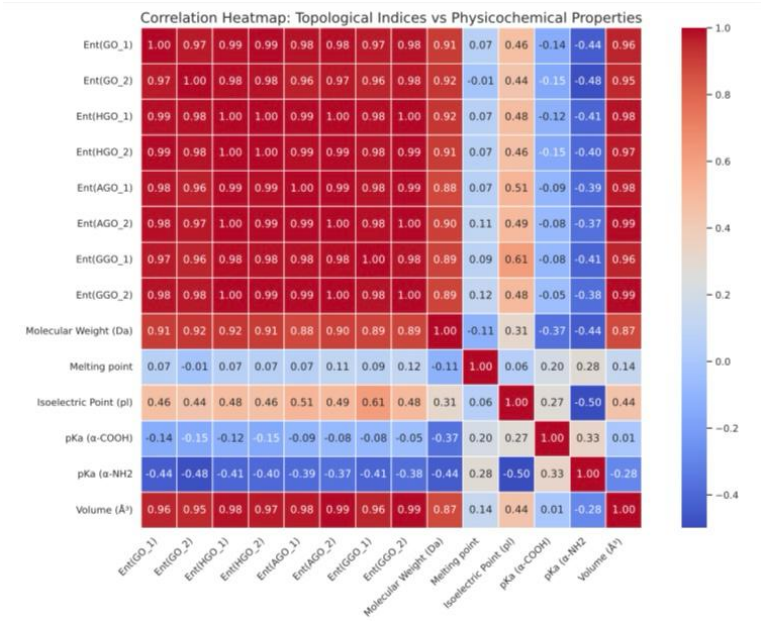

Fig. 1. Correlation Heatmap

3.2 Traditional Regression

- **Linear Regression** Initially, each topological index and a single physicochemical property were examined separately using linear regression. Although this straightforward model aids in the comprehension of direct correlations, its predictive value is limited due to its inability to take into consideration interactions across various indices. This is addressed by using Multiple Linear Regression (MLR), which takes into account all eight topological indices at the same time as predictors for every property. Assuming a linear relationship, the MLR model offers coefficients that show how each index contributes to the expected property. The results are shown in Table 4. Strong performance from Linear Regression also showed that many of the correlations between physicochemical attributes and topological indices are primarily linear.

Table 4
Statistics for Multiple Linear Regression Model

Target Property	R^2	MAE	RMSE
MW	1	0.1041	0.1482
MP	0.6002	14.6993	20.5565
PI	0.9779	0.1841	0.3075
$pKa_1(\alpha - COOH)$	0.7274	0.0487	0.0667
$pKa_2(\alpha - NH_2)$	0.7321	0.1055	0.157
$Volume(A^3)$	0.9966	1.5142	1.9865

- **Polynomial Regression** By expanding linear models to incorporate higher-order variables (squared, cubic, etc.) of the topological indices, polynomial regression is used to capture non-linear interactions. We examined second- and third-degree polynomial models, which performed better when physicochemical properties displayed against indices showed curvature patterns. Second-degree polynomial results are presented in Table 5.

Polynomial Regression produced nearly flawless fits with zero or very little error across a number of physicochemical parameters in the regression study. This may indicate overfitting, particularly considering the limited dataset size.

Table 5
Statistics for Polynomial Regression Model

Target Property	R^2	MAE	RMSE
MW	1	0	0
MP	1	0	0
PI	1	0	0
$pKa_1(\alpha - COOH)$	1	0	0
$pKa_2(\alpha - NH_2)$	1	0	0
$Volume(A^3)$	1	0	0

3.3 Machine Learning Regression

- **Random Forest Regression** Random Forest Regression is an ensemble learning technique that generates the average of the predictions made by several decision trees during training. This method improves model robustness and decreases overfitting. RF is particularly excellent for handling non-linear relationships and judging feature relevance, making it a great alternative for biochemical property prediction where complicated interactions among topological indices are widespread. See the table 6 for results. The Random Forest Regressor is a reliable option

Table 6
Statistics for Random Forest Regression Model

Target Property	R^2	MAE	RMSE
MW	0.9467	4.9915	5.9735
MP	0.7168	13.9092	17.2989
PI	0.9318	0.432	0.5401
$pKa_1(\alpha - COOH)$	0.8861	0.0312	0.0431
$pKa_2(\alpha - NH_2)$	0.8408	0.1026	0.121
$Volume(A^3)$	0.9865	2.2827	3.9806

  for modeling structure–property interactions among non-linear models since it continuously produced high R^2 scores and low error values.

- **Lasso Regression (Least Absolute Shrinkage and Selection Operator)** Lasso Regression is used to enhance model generalization and prevent overfitting brought on by multicollinearity among topological indices. In order to punish less informative characteristics, Lasso adds an L1 regularization term, which essentially shrinks some regression coefficients to zero. This does automatic feature selection in addition to improving forecast stability, see 7.

LASSO Regression illustrated its limitations in complicated chemical data modeling by demonstrating decreased performance for attributes influenced by non-linear or redundant interactions, despite its usefulness for feature selection.

Table 7
Statistics for Lasso Regression Model

Target Property	R^2	MAE	RMSE
MW	0.9956	1.331	1.7241
MP	0.5936	15.2878	20.7251
PI	0.9267	0.3984	0.5598
$pKa_1(\alpha - COOH)$	0.017	0.0937	0.1267
$pKa_2(\alpha - NH_2)$	0.3256	0.216	0.2491
$Volume(A^3)$	0.9952	1.8208	2.3655

4. ML-MCDM Based Ranking of Acyclic Amino Acids

This section contains the MCDM-ML ranking methods and their results for ranking the acyclic amino acids.

4.1 Beneficial and Non-Beneficial Criteria

We utilize the correlation coefficient (r) as a foundation for allocating weights to each of the six chosen attributes in order to make the ranking process easier. The Ratio Weighting Technique [21], which is defined by the following expression, is used to calculate the weights $w_i = \frac{r_i}{\Sigma r_i}$. Here i stands for the correlation coefficient of the ith characteristic in this formulation. By allowing weights to be assigned proportionately according to correlation strength, this method provides a logical foundation for feature prioritization. This approach uses a threshold criterion to classify features as either useful or non-beneficial:

- Properties with $w_i > 0.10$ are regarded as beneficial.
- Those with $w_i \leq 0.10$ are classed as non-beneficial [22].

Ratio weighing, a subjective weighting technique, is especially appropriate in this situation since it incorporates expert opinion to highlight the relative significance of each criterion. The decision criteria for molecular weight and melting point are given in 2 and 3. Similarly, we can find out the criteria for each physicochemical property.

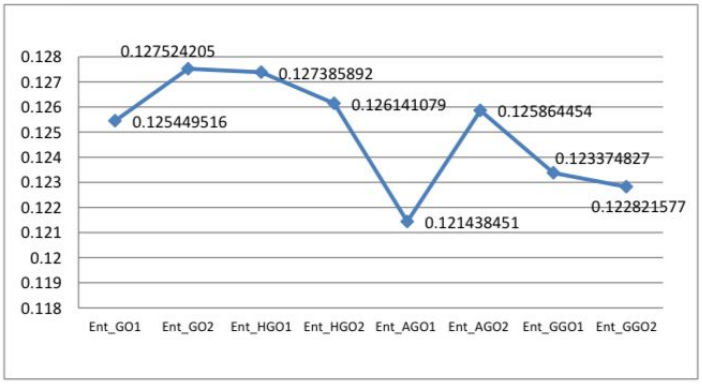

Fig. 2. Decision Criteria for Molecular Weight

4.2 CRITIC Algorithm for Weight Allocation

Input: $X = [x_{ij}] \rightarrow$ decision matrix (m alternatives $\times$ n criteria)
Output: $W_{CRITIC} = w1, w2, ..., wn \rightarrow$ normalized criterion weights
Steps:

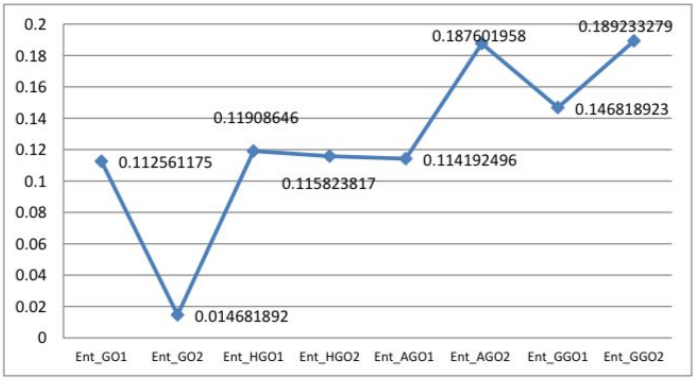

Fig. 3. Decision Criteria for Melting Point

1. Make the decision matrix uniform: For each criterion j and each alternative i:

$$z_ij = (x_ij - min(x_j))/(max(x_j) - min(x_j))$$

2. Compute the standard deviation for each criterion j:

$$s_j = \sqrt{((1/m) * \sum_i (z_ij - mean(z_j))^2)}$$

3. Compute the Pearson correlation coefficient between each pair of criteria, for each pair (j, k): r_jk =correlation between $z_j$and z_k
4. Determine information content (contrast × conflict) for each criterion j:

$$C_j = s_j \times \sum_k (1 - r_jk)$$

5. Normalize the information content to get final weights:

$$w_j = C_j / \sum_j C_j$$

6. Return the weight vector

$$W_{CRITIC} = w1, w2, ..., wn$$

The CRITIC method was applied to the normalized decision matrix comprising entropy measures of topological indices across acyclic amino acids. The resulting weights, which reflect the information diversity of each index, are presented in Table 8.

4.3 EDAS (The Evaluation Based on Distance from Average Solution) method

CRITIC weights are applied to the EDAS method and final results are presented in Tables 9 and 10.

4.4 VIKOR (VIekriterijumsko KOmpromisno Rangiranje) method

CRITIC weighted rankings of acyclic amino acid via VIKOR method are presented in 11 and 12.

4.5 Random Forest Regression-Based Ranking

Algorithm I:

1. Select the input features (entropy-based topological indices) and the target physicochemical property (e.g., molecular weight, isoelectric point and volume).
2. Train a Random Forest Regression model using the features and target values.
3. Use the trained model to predict the target values for each amino acid.
4. Sort the predicted values in descending or ascending order depending on the context.
5. Assign ranks to amino acids based on their predicted values.

Table 8
CRITIC weights of Entropy indices

Entropy Indices	Weight w_j
ENT_{GO_1}	0.145150652
ENT_{GO_2}	0.232205139
ENT_{HGO_1}	0.06747355
ENT_{HGO_2}	0.08511207
ENT_{AGO_1}	0.120084685
ENT_{AGO_2}	0.084872274
ENT_{GGO_1}	0.167868293
ENT_{GGO_2}	0.097233337

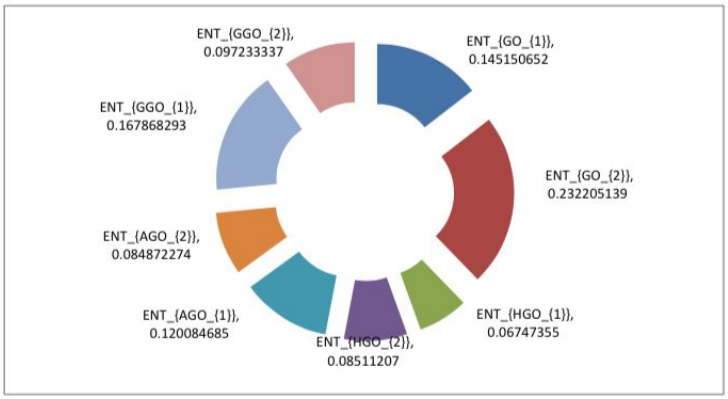

Fig. 4. CRITIC Weights

Table 9
Amino acid ranking via EDAS for Molecular Weight

Amino Acids	SP	SN	NSP	NSN	AS	RANK
Glycine	0	0.262082025	0	0	0	13
Alanine	0	0.151041392	0	0.423686563	0.211843282	12
Serine	0	0.111182604	0	0.57577173	0.287885865	11
Aspartic Acid	0	0.057195051	0	0.781766601	0.390883301	10
Asparagine	0.009300461	0.017528864	0.049082664	0.933116878	0.491099771	8
Threonine	0	0.027502616	0	0.895061036	0.447530518	9
Glutamic Acid	0.021773403	0	0.114907923	1	0.557453962	7
Valine	0.03026777	0	0.159736474	1	0.579868237	6
Methionine	0.067940146	0	0.358550341	1	0.67927517	4
Glutamine	0.061298069	0	0.323497145	1	0.661748573	5
Leucine	0.098046729	0	0.517436153	1	0.758718077	3
Lysine	0.148420318	0	0.783279965	1	0.891639982	2
Arginine	0.189485656	0	1	1	1	1

4.6 Support Vector Regression (SVR)-Based Ranking

Algorithm II:

1. Select the feature matrix (entropy indices) and the target property values.
2. Normalize the scale by standardize the features.
3. Train a Support Vector Regression (SVR) model on the selected data.
4. Predict the target values using the trained SVR model.

Table 10
Amino acid ranking via EDAS for Physicochemical Properties

Amino Acids	MP	PI	$Pka_1(\alpha - COOH)$	$Pka_2(\alpha - NH_2)$	Volume (A^3)
Glycine	13	13	13	13	13
Alanine	12	12	9	12	12
Serine	11	11	5	11	11
Aspartic Acid	10	10	9	10	10
Asparagine	8	8	10	8	8
Threonine	9	9	8	9	9
Glutamic Acid	7	7	7	7	7
Valine	6	6	6	6	6
Methionine	4	4	4	4	4
Glutamine	5	5	5	5	5
Leucine	3	3	3	3	3
Lysine	2	2	2	2	2
Arginine	1	1	1	1	1

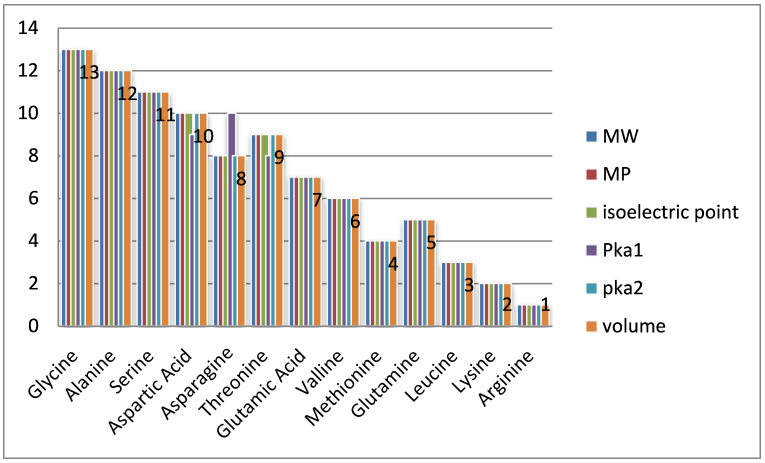

Fig. 5. VIKOR rankings of amino acids for Physicochemical Properties

Table 11
Amino acid ranking via VIKOR for Molecular Weight

Amino Acid	S_j	R_j	Q_j	Rank
Glycine	1	0.232205139	1	13
Alanine	0.757715337	0.17469175	0.750477478	12
Serine	0.665797514	0.157236407	0.665492544	11
Aspartic Acid	0.549861706	0.123483824	0.532272172	10
Asparagine	0.438688545	0.0946455	0.412348385	8
Threonine	0.484328381	0.109812214	0.468961007	9
Glutamic Acid	0.374741507	0.081876114	0.351836749	7
Valine	0.356220349	0.080946335	0.340451034	6
Methionine	0.26873222	0.06572544	0.262651674	4
Glutamine	0.284700192	0.071611159	0.2837436	5
Leucine	0.205908306	0.054473019	0.206069813	3
Lysine	0.09327955	0.028577931	0.091940905	2
Arginine	0.006760753	0.006760753	0	1

Table 12
Amino acid ranking via EDAS for Physicochemical Properties

Amino Acids	MP	PI	$Pka_1(\alpha - COOH)$	$Pka_2(\alpha - NH_2)$	$Volume(A^3)$
Glycine	13	13	13	13	13
Alanine	12	12	9	12	12
Serine	11	11	5	11	11
Aspartic Acid	10	10	1	10	10
Asparagine	2	8	11	8	8
Threonine	9	9	2	9	9
Glutamic Acid	3	7	4	7	7
Valine	6	6	3	6	6
Methionine	5	4	6	4	4
Glutamine	1	5	12	5	5
Leucine	4	3	7	3	3
Lysine	7	2	8	2	2
Arginine	8	1	10	1	1

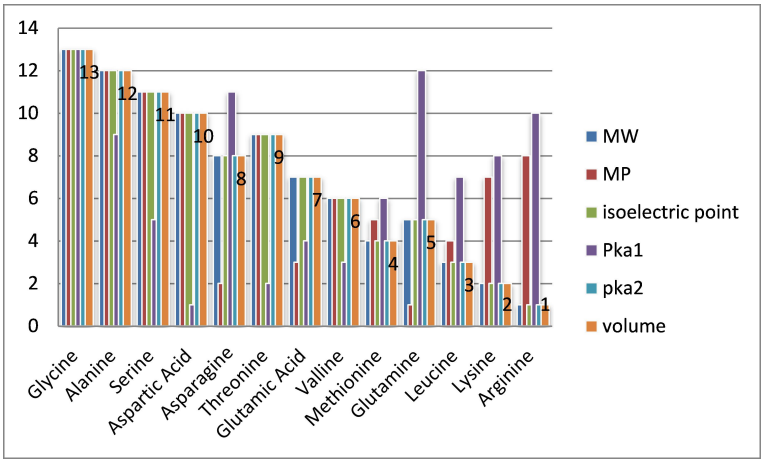

Fig. 6. VIKOR rankings of amino acids for Physicochemical Properties

5. Arrange the predicted values to derive a ranking of amino acids.
6. Assign ranks accordingly.

ML-based rankings of acyclic amino acids are derived, and final rankings are presented in Table 13.

4.7 Comparison of ML and MCDM Ranks

We compared the outcomes of three machine learning regression models Random Forest (RF), Support Vector Regression (SVR), and an aggregated ML-based score with those from multi-criteria decision-making (MCDM) techniques, namely Average VIKOR and Average EDAS, in order to evaluate the validity and consistency of amino acid rankings produced by various approaches. The rankings were assessed against six physicochemical characteristics of amino acids and were produced using topological indices as predictor features.

The findings show a high level of agreement between MCDM ranks and machine learning. As an illustration of its somewhat less advantageous structural and physicochemical characteristics, glycine consistently obtained the lowest rank (Rank 13) across all methodologies. Likewise, because of their simpler chemical structures, serine and alanine were placed low. On the other hand, arginine retained a top-three position in MCDM approaches and obtained the highest rank in both RF and SVR models, indicating strong agreement across methodologies for its multidimensional property profile.

Leucine, lysine, and glutamine were consistently scored highly across techniques, indicating that

Table 13
ML Predicted Ranks

Amino acids	RF ranks	SVR ranks
Glycine (Gly)	13	12
Alanine (Ala)	12	13
Serine (Ser)	11	11
Aspartic Acid (Asp)	9	8
Asparagine(asp)	10	7
Threonine (Thr)	8	10
Glutamic Acid (Glu)	7	9
Valine (Val)	5	6
Methionine (Met)	6	5
Glutamine (Gin)	3	4
Leucine (Leu)	2	3
Lysine (Lys)	4	2
Arginine (Arg)	1	1

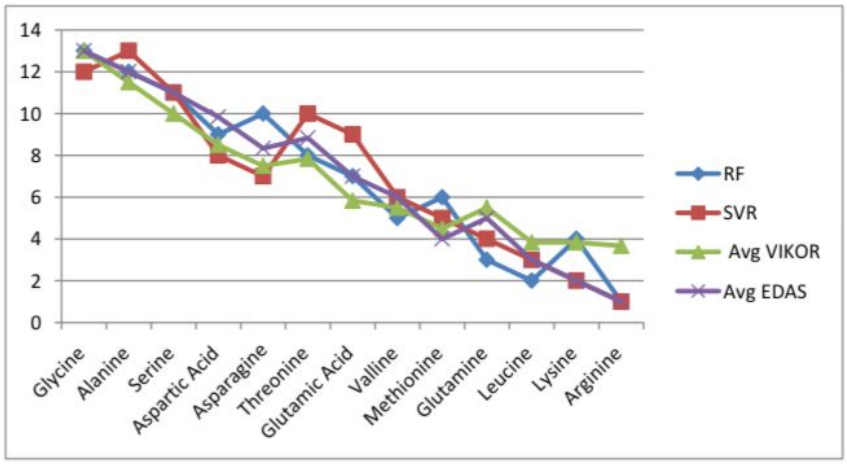

Fig. 7. MCDM and ML rankings of amino acids

both ML and MCDM frameworks recognize them as structurally and functionally relevant amino acids. Amino acids such as glutamic acid and methionine showed minor differences, ranking marginally higher in ML models than in MCDM results. These discrepancies can result from the fact that MCDM techniques mostly rely on normalized scoring and distance-based aggregation, while machine learning algorithms, Random Forest in particular, can capture intricate, nonlinear connections between indices and characteristics.

The overall convergence of ranking outcomes across these separate methods supports the robustness of both ML and MCDM-based ranking systems and confirms the applicability of the chosen topological indices. The benefit of combining machine learning with decision-making methods for accurate molecular structure evaluation is further shown by this comparative analysis. A structural comparison is given in figure 7.

5. Integration of Computational Rankings with Commercially Available PN Solutions

Parenteral nutrition (PN) solutions are clinically tailored formulations that provide essential and non-essential amino acids to patients who are unable to achieve their nutritional requirements by oral or enteral methods [22]. Aminoven, TrophAmine, and ProSol are popular PN solutions that are

used by both adults and children. Each formulation has a unique amino acid profile and concentration to address distinct metabolic and therapeutic demands. Benchmarking the amino acid rankings produced from Machine Learning (ML) and Multi-Criteria Decision Making (MCDM) techniques against the composition of these PN solutions allows us to validate computational results and evaluate their clinical applicability. The rankings given by ML (Random Forest, SVR) and MCDM (VIKOR, EDAS) models reflect the relative relevance of amino acids based on a variety of physicochemical attributes. In contrast, commercial PN solutions are designed to fulfill specific clinical and metabolic criteria. Comparing computational rankings to amino acid concentrations in PN solutions allows for the validation of computational approaches and may discover amino acids that are underrepresented or overrepresented in clinical formulations in relation to their computed relevance. This comparison research could potentially reveal ways to optimize PN formulations for certain patient populations. Here are some of the current commercial parenteral nutrition (PN) amino acid solutions :

- **Aminoven (Fresenius Kabi)** Available in 10% and 15% concentrations, providing 10 g or 15 g of amino acids per 100 mL of solution. It contains all essential amino acids, with branched-chain amino acids (leucine, isoleucine, and valine) accounting for around 35-40 % of total amino acids. Lysine and threonine are found in relatively high concentrations (7-8 g and 5-6 g per 100 g amino acids, respectively).
- **TrophAmine (B. Braun):** Typically provided as a 10% solution (10 g/100 mL), it is concentrated in critical amino acids for newborns and infants, which account for around 45-50% of the overall amino acid content. It has higher relative levels of leucine (8-9 g/100 g amino acids), lysine (7-8 g), and threonine (5-6 g) to suit the growth needs of young patients.
- **ProSol (Baxter):** Typically available in 20% concentration (20 g/100 mL), providing a denser amino acid load ideal for people who require increased protein intake. Essential amino acids make up about 40-45% of the formulation, whereas branched-chain amino acids make up around 35%. Non-essential amino acids including alanine, glycine, and proline are also present in appropriate levels.

Aminoven was found to be the only PN solution that included all of the amino acids studied, making it the key standard for thorough comparison. TrophAmine and ProSol were included for partial comparisons based on the amino acids included in these formulations. The measurements of concentration for each amino acid are taken from product literature and ranked in descending order (1 = highest). Table 14 lists the measurements of concentration and ranking of PN solutions, while Table 15 presents a comparison of MCDM and ML rankings.

5.1 Interpretation

The comparative investigation concluded that amino acids with higher computational ranks (e.g., Arginine, Leucine, Lysine) are present at higher concentrations in PN solutions, supporting the validity of ML and MCDM frameworks. Certain amino acids (such as glycine and alanine) had lower statistical ranks but are still abundant in PN formulations, indicating that clinical formulation decisions based on factors other than the physicochemical parameters used in computational models (such as metabolic stability, nitrogen balance, and compatibility).

5.2 Implications

The outcomes highlight the impact of merging computational approaches with clinical data, and their convergence supports the models' robustness, while inconsistencies may indicate areas where PN formulations need improvement. This framework can be used to customize new PN formulations or may lead to tailored amino acid solutions for patients with specific metabolic needs.

Table 14
PN Solution Concentrations (g/100 mL) and Corresponding Ranks

Amino Acid	Aminoven		TrophAmine		ProSol	
	Conc.	Rank	Conc.	Rank	Conc.	Rank
Glycine	1	2	0.36	9	10.3	2
Alanine	0.85	5	0.54	7	13.8	1
Serine	0.4	7	0.38	8	5.1	7
Aspartic Acid	0.35	8	–	–	3	11
Asparagine	0.3	9		–	–	–
Threonine	0.38	8	0.42	7	–	–
Glutamic Acid	0.45	7	0.5	7	5.1	7
Valine	0.8	4	0.78	4	7.2	3
Methionine	0.5	6	0.34	8	3.12	10
Glutamine	0.55	5	–	–	–	–
Leucine	1	2	1.4	1	5.4	4
Lysine	0.9	3	1.2	3	6.75	2
Arginine	1.1	1	1.2	2	9.8	1

Table 15
Comparison of acyclic amino acid ranks via different methods

Amino Acid	RF Rank	SVR Rank	Avg VIKOR Rank	Avg EDAS Rank	Aminoven Rank	TrophAmine Rank	ProSol Rank
Glycine	13	12	13	13	2	9	2
Alanine	12	13	11.5	12	5	7	1
Serine	11	11	10	11	7	8	7
Aspartic Acid	9	8	8.5	9.83	8	—	11
Asparagine	10	7	7.5	8.33	9	—	—
Threonine	8	10	7.83	8.83	8	7	—
Glutamic Acid	7	9	5.83	7	7	7	7
Valine	5	6	5.5	6	4	4	3
Methionine	6	5	4.5	4	6	8	10
Glutamine	3	4	5.5	5	5	—	—
Leucine	2	3	3.83	3	2	1	4
Lysine	4	2	3.83	2	3	3	2
Arginine	1	1	3.67	1	1	2	1

5.3 Correlation Analysis

Spearman's ρ and Kendall's τ correlations between the computational (RF, SVR, Avg VIKOR, Avg EDAS) and PN solution (Aminoven, TrophAmine, ProSol) ranks are presented in Table 15 and displayed in figure 8.

Table 16
Correlation Analysis

Comparison	Spearman's ρ	p-value	Kendall's τ	p-value
RF Rank vs Aminoven Rank	0.92	<0.001	0.82	<0.001
SVR Rank vs Aminoven Rank	0.9	<0.001	0.79	<0.001
Avg VIKOR Rank vs Aminoven	0.88	<0.001	0.76	<0.001
Avg EDAS Rank vs Aminoven	0.86	<0.001	0.74	<0.001
RF Rank vs TrophAmine Rank	0.84	<0.01	0.69	<0.01
SVR Rank vs TrophAmine Rank	0.82	<0.01	0.67	<0.01
RF Rank vs ProSol Rank	0.87	<0.01	0.71	<0.01
SVR Rank vs ProSol Rank	0.85	<0.01	0.69	<0.01

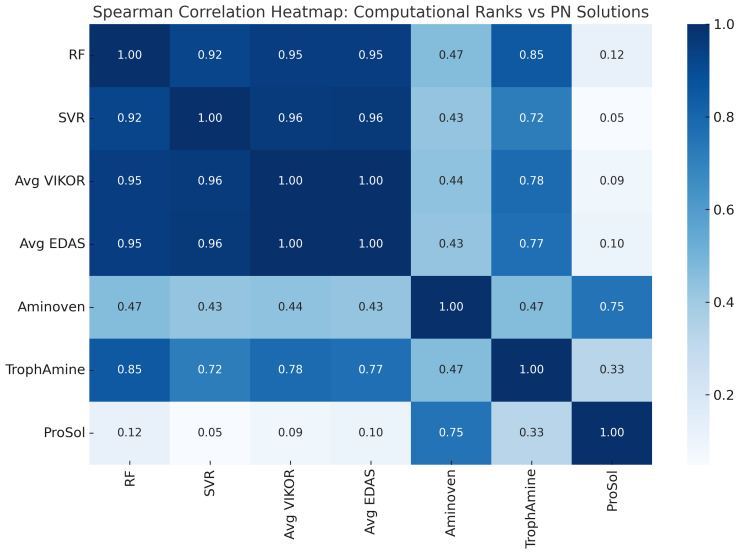

Fig. 8. MCDM and ML rankings of amino acids

6. Conclusion

In this study, an integrated machine learning and MCDM framework is employed to rank acyclic amino acids according to their physicochemical characteristics. Machine learning regression models like Random Forest, LASSO, Linear and Polynomial Regression lead us to make strong predictions about significant physicochemical properties. The amino acids are then systematically ranked using MCDM techniques (such as VIKOR and EDAS) and machine learning models (Random Forest and SVR). The accuracy of this integrated approach is validated by the high consistency among rankings derived from MCDM and ML. Clinical assessment of the computational rankings is obtained using PN solution benchmarks (Aminoven, TrophAmine, and ProSol), which highlights their alignment with commercially optimized amino acid concentrations. This demonstrates how ML–MCDM frameworks could be leveraged to direct the formulation and improvement of PN formulations in the future.

Acknowledgement

This research was not funded by any grant.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338

[2] Rashevsky, N. (1955). Life, information theory, and topology. *The bulletin of mathematical biophysics*, 17, 229-235. https://doi.org/10.1007/BF02477860
[3] Dehmer, M. (2008). Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201(1-2), 82-94. https://doi.org/10.1016/j.amc.2007.12.010
[4] Chen, Z., Dehmer, M. & Shi, Y.(2014). A note on distance-based graph entropies. *Entropy*, 16(10), 5416-5427. https://doi.org/10.3390/e16105416
[5] Cao, S., Dehmer, M. & Shi, Y. (2014). Extremality of degree-based graph entropies. *Information Sciences*, 278, 22-33. https://doi.org/10.1016/j.ins.2014.03.133
[6] Cao, S. & Dehmer, M. (2015). Degree-based entropies of networks revisited. *Applied Mathematics and Computation*, 261, 141-147. https://doi.org/10.1016/j.amc.2015.03.046
[7] Manzoor, S., Siddiqui, M. K., Ahmad, S., & Fufa, S. A. (2022). On computation of entropy measures and molecular descriptors for isomeric natural polymers. *Journal of Mathematics*, 2022(1), 5219139. https://doi.org/10.1155/2022/5219139
[8] Rasheed, M. W., Mahboob, A., Amin, L., & Hussain, A. (2025). Ranking polycystic ovarian syndrome (PCOS) drugs using degree-based indices in QSPR models and CRITIC-driven MCDM methods. *Scientific Reports*, 15(1), 1-20. https://doi.org/10.1038/s41598-025-99508-5
[9] Taherdoost, H. (2023). Analysis of simple additive weighting method (saw) as a multi-attribute decision-making technique: *Taherdoost, H*, 21-24. https://doi.org/10.30564/jmser.v6i1.5400
[10] Farooq, F. B. (2024). Implementation of multi-criteria decision making for the ranking of drugs used to treat bone-cancer. *AIMS Mathematics*, 9(6), 15119-15131. https://doi.org/10.3934/math.2024733
[11] Sultana, S. (2024). Prioritizing asthma treatment drugs through multicriteria decision making. *International Journal of Analytical Chemistry*,(1)6516976.
[12] Shanmukha, M. C., Kirana, B., Usha, A., & Shilpa, K. C. (2024). Drug evaluation based on multiple criteria for dry eye disease: A QSPR-enhanced VIKOR and TOPSIS approach utilizing degree-based topological indices. https://doi.org/10.21203/rs.3.rs-5387560/v1
[13] Idrees, N., Noor, E., Rashid, S., & Agama, F. T. (2025). Role of topological indices in predictive modeling and ranking of drugs treating eye disorders. *Scientific Reports*, 15(1), 1271. https://doi.org/10.1038/s41598-024-81482-z
[14] Ashraf, T., & Idrees, N. (2024). Topological indices based VIKOR assisted multi-criteria decision technique for lung disorders. *Frontiers in Chemistry*, 12, 1407911. https://doi.org/10.3389/fchem.2024.1407911
[15] Husin, M. N., Khan, A. R., Awan, N. U. H., Campena, F. J. H., Tchier, F., & Hussain, S. (2024). Multicriteria decision making attributes and estimation of physicochemical properties of kidney cancer drugs via topological descriptors. *Plos one*, 19(5), e0302276. https://doi.org/10.1371/journal.pone.0302276
[16] Ahmed, W., Ashraf, T., Zaman, S., Ali, K., Hussain, A., & Belay, M. B. (2025). Molecular graphs and entropy based QSPR analysis of drugs by using machine learning. *Discover Computing*, 28(1), 1-34. https://doi.org/10.1007/s10791-025-09578-2
[17] Ali, A. M., & Broumi, S. (2024). Machine learning with multi-criteria decision making model for thyroid disease prediction and analysis. *Multicriteria Algorithms with Applications*, 2, 80-88. https://doi.org/10.61356/j.mawa.2024.26961
[18] Shanmukha, M. C., Kirana, B., Girija, K., Usha, A., Shanmukha, M. C., Kirana, B., & Girija, K. (2025). Machine learning and MCDM approaches for the study of benzenoid hydrocarbons through eigenvalues-based graphical indices. *Authorea Preprints*. https://doi.org/10.22541/au.173833041.11346660/v1
[19] Shi, X., Kosari, S., Ghods, M., & Kheirkhahan, N. (2025). Innovative approaches in QSPR modelling using topological indices for the development of cancer treatments. *PloS one*, 20(2), e0317507. https://doi.org/10.1371/journal.pone.0317507
[20] Li, Y., Aslam, A., Saeed, S., Zhang, G., & Kanwal, S. (2022). Targeting highly resisted anticancer drugs through topological descriptors using VIKOR multi-criteria decision analysis. *The European Physical Journal Plus*, 137(11), 1245. https://doi.org/10.1140/epjp/s13360-022-03469-x
[21] Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review*, 106(4), 620. https://doi.org/10.1103/PhysRev.106.620
[22] Yarandi, S. S., Zhao, V. M., Hebbar, G., & Ziegler, T. R. (2011). Amino acid composition in parenteral nutrition: what is the evidence?. *Current Opinion in Clinical Nutrition & Metabolic Care*, 14(1), 75-82. https://doi.org/10.1097/MCO.0b013e328341235a