

Nutritional Valorization of Oilseed Cakes: A Comparative Study on Proximate Composition, Anti-Nutritional Factors, and AI-Driven Optimization of ANF Reduction Strategies

Betcey Teresa Tomy¹, Narinder Kaur^{*,1}, Yasir afaq²

¹Department of Food Technology and Nutrition,
Lovely Professional University, Jalandhar-144411, India

²Department of computer Science and Engineering,
SRM University-Amravati, Andhra Pradesh, India

*Corresponding author: Dr. Narinder Kaur, (kaurnarinder88@gmail.com)

Abstract

Oilseed cakes are high-protein by-products of the oil extraction industry with considerable potential for food fortification and nutrition. They are generally rich in its nutritional composition and are agriculturally and industrially valuable by-products. Three oilseed cakes namely soybean (45-50%), sesame (45-50%), and groundnut (40-50%) were selected due to their higher protein content. However, the presence of potent anti-nutritional factors (ANFs) such as trypsin inhibitors, phytates, and tannins limit their bio-utilization by impairing digestibility and solubility. The research paper presents a comparative analysis of the nutritional composition of these three oilcakes and assesses their specific ANF profiles. Furthermore, this study proposes a hybrid computational framework utilizing Artificial Neural Networks (ANN) and Genetic Algorithms (GA) to optimize thermal and chemical pre-treatments. The aim is to attain maximum ANF reduction while maintaining the functional integrity of the proteins for subsequent enzymatic hydrolysis. A balanced dataset of 300 samples was generated, which was computational with 100 samples for each oilcake. Each sample represented a unique combination of treatments like thermal, chemical and enzymatic. The input variables consist of Temperature (100–121°C), Time (5–20 min), pH (5.5–7.0), Ethanol percent (20–60%), and Enzyme(phytase) percent (0.5–1.5%). Protein Solubility, Residual Phytates, Residual Tannins, and Residual Trypsin Inhibitors were the output variables. Soybean exhibited the highest mean protein solubility (97.19%) among the three. It was achieved at slightly lower temperatures (104.94°C) and slightly higher pH values (5.97), alongside very effective (reduced to 0.005%) reduction of residual tannins. Sesame required longer processing time (7.04 min) and higher ethanol concentrations (30.92%), indicating a need for more intensive treatment to effectively reduce its typically higher phytate and tannin levels. For groundnut, optimal conditions involved moderate temperatures (107°C) and ethanol percentages (20%), suggesting an efficient process for achieving high protein solubility and balanced ANF reduction.

Keywords: Oilcakes- Soybean, Sesame, Groundnut, Anti-Nutritional Factors (ANF), Artificial Neural Networks (ANN), Random Forest, Genetic Algorithms (GA)

How to cite this article: Betcey Teresa Tomy¹ Narinder Kaur^{*} 1 Yasir afaq². Nutritional Valorization of Oilseed Cakes: A Comparative Study on Proximate Composition, Anti-Nutritional Factors, and AI-Driven Optimization of ANF Reduction Strategies. *Int J Drug Deliv Technol.* 2026;16(81s):1515-1525. DOI: 10.25258/ijddt.16.81s.185.

1. Introduction

Oilcakes are the residue obtained after extraction of oil from oilseeds. Oilcakes like oilseeds are rich in its nutritional composition. They are coarse and contain proteins, minerals and are thus agriculturally and industrially valuable by-products (Vichare, S. A., & Morya, S. 2024). Generally, oilseeds have good protein content of 45 to 60%. The protein content varies according to oilseed and processing method. Ground nut cake has 45 to 55% (Singh et al., 2022), Soybean cake has 45 to 50% (Popovic et al., 2020), sesame cake has 40 to 50% (Nouska et al., 2024) sunflower cake has 40-50%, cotton cake has 45% (Vichare, S. A., & Morya, S. 2024).

Oil cakes are an excellent source of protein and energy for animals and are therefore used as feed for

poultry, pigs, and cattle. They can be added to pelleted feeds to enhance handling, processing, and digestibility (Rakita et al., 2022). Its entry into food industry is often limited. However, they can be incorporated into fortified flour mixtures, breakfast cereals, snacks, and beverages to enhance protein, fibre, and micronutrient levels. They can also be used to improve the texture and mouth feel of baked foods and meat substitutes (Kaur et al., 2022). Protein energy malnutrition is one of the major problems our country is facing and this is the reason we are unable to attain sustainable development goal 2 'Zero hunger' (Katoch 2024). This assists in alleviating malnutrition and enhancing community health benefits.

Oil cakes can serve as a substrate for the fermentation of a range of products, such as enzymes, antibiotics, mushrooms, and vitamins. This enables the manufacture of low-cost, value-added products. They can also be utilized in the manufacture of biofuels, surfactants, and natural pigments (Ramachandran et al., 2007). Oilcakes may serve as a natural texturizer, binder, or emulsifier in foods. They can also find application in the manufacture of protein isolates and concentrates for food (Zhang et al., 2023).

The global demand for plant-based proteins has shifted focus toward agricultural by-products like oilcakes. But often they are locked in it with limited digestibility and solubility due to the presence of anti-nutritional factors (ANF) (Singh et al., 2022.). Despite their nutritional density, the presence of ANFs creates a "nutritional barrier"

Antinutritional factors in oilcake

ANFs are naturally occurring physiologically active substances that disrupt the absorption, digestion, and metabolic use of vital nutrients in plant-based meals. Plants produce these secondary metabolites as a defense mechanism against herbivores, insects, diseases, and unfavorable growing circumstances because they are unable to flee predators (Popova, A., & Mihaylova, D. (2019). When ingested by monogastric animals or humans, they serve as a "nutritional barrier." Antinutritional compounds normally found in oilcakes are trypsin inhibitors, phytates, tannins, along with others like lipoxigenase, oligosaccharides, lectins, and saponins (Salim et al., 2023)

Heat-Labile ANFs can be easily transformed and inactivated by heat processing, such as roasting, autoclaving, or extrusion. Examples include trypsin inhibitors, lectins, and goitrogens. Heat-Stable ANFs can resist heat and need chemical treatments, fermentation, enzymatic hydrolysis, or physical extraction techniques like soaking and dehulling for removal. Examples include phytic acid, tannins, and saponins (Arsov et al., 2024)

ANF reduction:

For efficient utilization of oilcakes in the food industry, it is necessary to remove these ANFs without affecting the protein compounds in them. Although there are several methods like, thermal processing, chemical treatment, enzymatic hydrolysis, soaking, fermentation, roasting, boiling to remove ANFs, it is important to find which method at what processing parameter removes them the most efficiently. This is where Artificial intelligence (AI) driven optimization of the ANF reduction strategy is used. The optimization was executed using a Genetic Algorithm (GA) driven by Random Forest and Artificial Neural Networks (ANN) models. The

process was aimed at enhancing protein solubility and minimizing anti-nutritional factors (ANFs) in Groundnut, Sesame, and Soybean oilcakes.

Materials and Methodology

Cold pressed oilcakes namely sesame, soyabean and groundnut were collected after the extraction of oil from Dharapuram region of Tirupur district of Tamil Nadu.

The nutritional composition of oilcakes is tested using the following methods.

- Drying the sample to constant weight to know the moisture content, (AOAC 2005)
- Kjeldahl method to know the crude protein, (AOAC 2005)
- Soxhlet extraction to know the fat content, (AOAC 2005)
- Incinerating the sample and measuring the residue to know the ash content, (AOAC 2005)
- Calculating the difference after determining other components to know the carbohydrate content,
- Digestion with acid/ alkali to determine the crude fibre,
- Atomic absorption spectrophotometry (AAS) to know the concentration of Ca, Mg, Fe, Zn, Cu, Mn

The anti-nutritional factors were determined using the following method:

- Phytic Acid: Measured using Wade's Reagent Method
- Tannins: Determined using the Folin-Denis reagent and tannic acid as a standard, measured at 680 nm. (Nair et al., 2015), (Polshettiwar et al., 2007)
- Trypsin Inhibitors (T.I.): Measured using the AOCS method (2009).

ANF reduction and Protein solubility enhancement

Dataset Development and Experimental Design

To enhance protein solubility and reduce anti-nutritional factors (ANFs), the study developed an artificial intelligence-based multi-objective optimization framework in Soybean, Sesame, and Groundnut oilcakes. A balanced dataset of 300 samples was generated, which was computational with 100 samples for each oilcake. Each sample represented a unique combination of treatments like thermal, chemical and enzymatic.

Temperature C (100–121°C), Time (5–20 min), pH (5.5–7.0), Ethanol percent (20–60%), and Enzyme percent (0.5–1.5%) were the input variables. The output variables included Protein Solubility, Residual Phytates, Residual Tannins, and Residual Trypsin Inhibitors. The dataset was generated to simulate food processing conditions where moderate thermal

and enzymatic treatments enhance protein functionality and reduce ANFs with least protein denaturation.

Data Preprocessing

Python-based Pandas and NumPy libraries were used for data preprocessing. Oilcake types were converted into numerical representations using Label Encoding to enable machine learning analysis. The confirmed dataset was split into training and testing subsets using an 80:20 train-test split ratio for predictive modelling and validation.

Random Forest and ANN Modeling

Separate development of Random Forest Regression and Artificial Neural Network (ANN) models were done for foreseeing Protein Solubility, Residual Phytates, Residual Tannins, and Residual Trypsin Inhibitors.

200 decision trees were used to constitute Random Forest models to enhance nonlinear learning capability and predictive stability. ANN models were configured using TensorFlow and Keras frameworks with fully connected dense neural architectures containing hidden layers with ReLU activation functions.

Mean Absolute Error (MAE), Root Mean Square Error (RMSE), coefficient of determination (R^2), and 5-fold cross-validation were used to assess model performance. To find the most suitable predictive framework for optimization of ANF reduction and protein solubility, comparative analysis between Random Forest and ANN models was held.

Multi-Objective Genetic Algorithm Optimization

Using the PyGAD library, the optimized machine learning models were combined into a multi-objective Genetic Algorithm (GA) framework. The optimization objective concurrently maximized Protein Solubility while minimizing Residual Phytates, Residual Tannins, and Residual Trypsin Inhibitors.

Independent oilcake-specific optimization was done for Soybean, Sesame, and Groundnut oilcakes to guarantee balanced comparative optimization and eliminate prejudice for a single type. The optimization process was held within experimentally constrained processing ranges to ensure correctness and validity.

SHAP Explainability Analysis

SHAP (SHapley Additive exPlanations) analysis was applied to enhance model interpretability and avoid black-box optimization limitations for the Random Forest Protein Solubility model. To compute the effect of thermal, chemical, and enzymatic variables on oilcake processing, SHAP summary plots, feature importance analysis, and dependence plots were generated

Comparative Visualization and Trade-Off Analysis

Comparative visualization and trade-off analyses were achieved using optimization plots, radar charts, and actual-versus-predicted visualizations to assess the balance between improvement of nutrients and minimization of ANF in all oilcake types.

Computational Environment

All evaluations were carried out in the Jupyter Notebook environment using Python programming language. Scikit-learn was used for Random Forest modeling, TensorFlow/Keras for ANN development, PyGAD for optimization, SHAP for explainability analysis, and Matplotlib/Seaborn for visualization and graphical interpretation.

3. Results and Discussion

Table 3.1: Nutritional composition in oilcakes (g/100g)

Nutritional Component	Soybean Oilcake	Sesame Oilcake	Groundnut Oilcake
Macronutrients			
Crude Protein	37.67±2.57	30.74±0.93	35.12±2.04
Carbohydrate	51.53±1.22	45.43±1.49	46.06±3.20
Fat	6.39±0.93	13.50±1.00	8.12±1.04
Crude Fibre	4.62±0.90	4.41±1.12	3.94±1.29
Ash Content	5.38±0.85	8.51±0.79	5.33±1.88
Moisture	6.40±0.93	6.29±0.71	6.26±0.84
Key Minerals			
Potassium (K)	2.62±0.89	1.63±0.74	1.64±0.70
Phosphorus (P)	0.98±0.22	0.97±0.42	0.67±0.21
Calcium (Ca)	0.62±0.26	1.93±0.49	0.45±0.37
Magnesium (Mg)	0.71±0.44	0.62±0.26	0.32±0.09
Iron (Fe)	0.018±0.002	0.021±0.005	0.06±0.07
Zinc (Zn)	0.005±0.002	0.013±0.001	0.008±0.002

Nutritional profile analysis of soybean, sesame, and groundnut oilcakes shows considerable variations in their macronutrient and mineral compositions.

Macronutrient Composition: Soybean oilcake has the highest crude protein (37.67±1.57%) and carbohydrate (51.53±1.22%) levels compared to

others. Groundnut oilcake has a similar but slightly lower protein content ($35.12 \pm 2.04\%$). These values are slightly lower than the values reported by Bamgbose, 2013 (38-42% for Soy bean oilcake and 35-40% for Groundnut oilcake). This is seen because of the extraction method. Cold pressed cakes show lower protein content when compared to solvent extracted cakes. In contrast, sesame oilcake has a much higher lipid concentration, measuring $13.50 \pm 1.00\%$, which is more than double that of soybean oilcake ($6.39 \pm 0.93\%$). This makes sesame oilcake the most energy-dense option. These values are comparable to the values reported by Das et al., 2015.

Sesame and soyabean oilcake have almost the same crude fiber value ($4.41 \pm 1.12\%$; $4.62 \pm 0.92\%$). Groundnut oilcake has the lowest fiber content ($3.06 \pm 0.03\%$), which may lead to better digestibility. Dehulling in groundnut is the reason for low fibre content. This trend is supported by Behara et al., 2013 and Latif et al., 2008. The moisture content is almost the same in all three oilcakes

Mineral and Ash Profile: Total mineral density, measured by ash content, is highest in sesame oilcake ($8.02 \pm 0.02\%$). These results are comparable to the values reported by Sunil et al., 2015. It contains $1.93 \pm 0.49\%$ calcium, more than soybean ($0.62 \pm 0.26\%$) and groundnut ($0.55 \pm 0.37\%$). Bamgbose, 2013 confirms that the calcium values are higher in sesame and soyabean compared to others.

However, soybean oilcake claims the highest potassium level ($2.62 \pm 0.89\%$), which is essential for intracellular electrolytes. Baisch et al., 2024 also recorded the same value. This level is nearly double that found in the other samples, along with magnesium content at ($0.71 \pm 0.44\%$). The concentration of iron is high ($0.06 \pm 0.07\%$) in groundnut oilcake. Nevertheless, sesame oilcake has the highest zinc concentration ($0.013 \pm 0.001\%$), which is double that in soybean oilcake ($0.005 \pm 0.003\%$). This makes sesame oilcake beneficial for enzymatic and immune formulations.

Soybean oilcake is ideal for diets requiring higher protein, carbohydrate, and potassium levels to support growth. Sesame oilcake is best for energy-dense needs and to address mineral deficiencies, particularly in calcium and zinc. Groundnut oilcake provides a balanced macronutrient profile with lower fiber content, making it favorable for easily digestible formulations.

Table 3.2: ANF in oilcakes (mg/g)

Anti-Nutritional Factor	Soybean	Sesame	Groundnut	Primary effect
-------------------------	---------	--------	-----------	----------------

Phytates	17.43 ± 0.26	33.5 ± 0.16	13.36 ± 0.20	Chelates essential minerals like Ca, Fe, Zn and cross-links with proteins, reducing their solubility and digestibility.
Trypsin Inhibitors	6.36 ± 0.24	0.26 ± 0.12	4.3 ± 0.14	Deactivates digestive protease enzymes like trypsin and chymotrypsin by directly binding to them
Tannins	3.5 ± 0.21	14.3 ± 0.16	3.5 ± 0.05	Inhibits the activity of various digestive enzymes by precipitating protein

Palange et al., 2025 reports values of phytates in soyabean between 10-20mg/g which includes the values ($17.43 \pm 0.26\text{mg/g}$) reported here. The values reported for Trypsin inhibitors and tannins in soyabean by Di et al., 2025 is 6-10mg/g and 2-4 mg/g which includes the values ($6.36 \pm 0.24\text{mg/g}$ and $3.5 \pm 0.21\text{mg/g}$) obtained here. The value reported by

Aniyam et al., 2025 for phytates in sesame is 30-50 mg/g which is comparable to the value (33.5 ± 0.16 mg/g) reported here. The values reported by Embaby, 2010 for trypsin inhibitors in sesame and groundnut oilcake are 0.1-0.5 mg/g and 2-5 mg/g which are comparable to the values (0.26 ± 0.12 mg/g and 4.3 ± 0.14 mg/g) obtained here.

Phytates (Phytic Acid):

Phytic acid binds to minerals like iron, zinc, and calcium, reducing their availability and making them harder to absorb. Soaking and boiling can decrease phytate levels by leaching and heat degradation (Lee et al., 2021). Fermentation boosts the breakdown of phytic acid, especially when using *Aspergillus oryzae* (Gamba et al., 2024). Enzymatic treatment involves adding phytase enzymes directly, which break down phytic acid. Sprouting and germination activate natural phytase enzymes, although this method is less convenient for oilcakes. Other methods, such as dehulling, milling, ion-exchange resins, and ultrafiltration, also lower phytate content in soy products (Lee et al., 2021).

Tannins:

Tannins are water-soluble compounds found in many plants. They have multiple hydroxyl groups that form stable bonds with proteins through hydrogen bonding and hydrophobic interactions. This binding changes the protein structure, causing it to clump together and reducing its solubility and extractability (Ozdalet al., 2013). Tannins also bind to and inhibit digestive enzymes like trypsin and chymotrypsin, as well as enzymes that break down carbohydrates, while forming complexes with substrate proteins (Gilani et al., 2005). Like phytic acid, tannins create insoluble complexes with essential metal ions, reducing their availability in the digestive system. Tannins are fairly heat-stable, so standard cooking methods like roasting do not fully remove them. Effective methods to reduce tannins include dehulling (Mugendi et al., 2010), chemical leaching using alkaline solutions or solvents (Khattab et al., 2009), or solid-state fermentation to break down polyphenol-protein complexes before enzymatic hydrolysis (Manjit et al., 2008).

Phytate and Tannin Load in Sesame:

Sesame oilcake has the highest levels of phytates and tannins. Its phytate concentration (33.5 ± 0.16) is more than double that of groundnut oilcake (13.36 ± 0.20) and nearly twice that of soybean (17.43 ± 0.26). The tannin level in sesame oilcake is also higher at 14.3 ± 0.16 , while soybean and groundnut oilcake both contain only 3.5. This combination of mineral-binding phytates and protein-precipitating tannins creates a unique nutritional challenge. Although sesame oilcake is rich in minerals, the body cannot use them effectively unless these compounds are

removed through heat treatment, soaking, or fermentation.

Trypsin Inhibitors:

Trypsin inhibitors block protein digestion by inhibiting proteolytic enzymes in the digestive system (Li et al., 2017). Heat treatment, such as roasting or steaming at around 100°C for 30 minutes, efficiently denatures these enzyme inhibitors. Extrusion cooking applies heat, pressure, and shear simultaneously to inactivate inhibitors. Desolventizing-toasting is used in oil extraction to eliminate solvents and reduce antinutrients. Fermentation with microbes, such as *Lactobacillus* spp., can also lower protease inhibitor levels (Sessa and Nelson, 1991). Soaking and adjusting the pH with acids or alkalis can break down inhibitors. Chemical agents like sodium metabisulfite and ascorbic acid have been used to inactivate these inhibitors. For example, sodium metabisulfite treatment resulted in over 85% inactivation of Kunitz and Bowman-Birk inhibitors (Li et al., 2017).

Protein Digestibility Challenges in Soybean:

While soybean oilcake has low tannin levels, it has the highest amount of trypsin inhibitors (6.36 ± 0.24), followed by groundnut oilcake (4.3 ± 0.14). Sesame oilcake contains very little trypsin inhibitors (0.26 ± 0.12). One major issue with trypsin inhibitors is that they impair protein digestion. However, they are sensitive to heat and can be removed effectively through cooking.

Overall Performance and Key Factors

The mean optimized parameters and their predicted outcomes for each oilcake type are summarized in table 3.3

Table 3.3: Optimized Parameters and Predicted Outcomes

Oil cake	A v g. T e m p (°C)	A v g. T e m p (°C)	A v g. p H	A v g. E t h a n o l (%)	A v g. E n z y m e (%)	A v g. P r o t e i n S o l u b i l i t y	A v g. R e s i d u a l P h y t a t e s	A v g. R e s i d u a l T a n n i n s	A v g. R e s i d u a l T r y p s i n I n h i b i t o r s
Soybean	104.94	6.34	5.97	26.96	0.99	97.19	7.91	0.05	3.11

Nutritional Valorization of Oilseed Cakes: A Comparative Study on Proximate Composition, Anti-Nutritional Factors, and AI-Driven Optimization of ANF Reduction Strategies

Sesame	106.1	7.4	5.74	30.92	0.92	95.31	10.36	1.34	0.62
Groundnut	107.0	5.0	6.93	20.0	0.67	96.32	9.65	4.77	0.71

Note: Values are rounded mean predictions from the optimization results.

Soybean: Exhibited the highest mean protein solubility (97.19%) among the three. It was achieved at slightly lower temperatures (104.94C) and slightly higher pH values (5.97), alongside very effective (reduced to 0.005%) reduction of residual tannins.

Sesame: Required slightly longer processing time (7.04 min) and higher ethanol concentrations (30.92%), indicating a need for more intensive treatment to effectively reduce its typically higher phytate and tannin levels.

Groundnut: Optimal conditions involved moderate temperatures (107C) and ethanol percentages (20%), suggesting an efficient process for achieving high protein solubility and balanced ANF reduction.

Model Performance Comparison

The predictive models (Random Forest, initial ANN, scaled ANN, and tuned ANN) were thoroughly evaluated. The comparison of MAE, RMSE, and R2 scores in table 3.4 validates the reliability of these models in guiding the optimization process. Notably, the tuned ANN for Protein Solubility showed performance comparable to the Random Forest model for this specific target, especially after feature scaling and hyperparameter optimization.

Table 3.4: Comprehensive Model Performance Comparison (MAE, RMSE, R2)

Model	ANN (Initial)	ANN (Scaled)	ANN (Tuned Solubility)	Random Forest
Protein Solubility	4.1799	7.3130	4.7030	0.7252
Residual Phytates	7.7002	1.0840	NaN	0.6327
Residual Tannins	4.4877	0.4885	NaN	0.1823

Residual TIA	0.5680	0.1531	NaN	0.1677
Protein Solubility	-6.9683	-25.2181	-10.0934	0.5050
Residual Phytates	-0.0478	0.9732	NaN	0.9901
Residual Tannins	-0.0256	0.9800	NaN	0.9970
Residual TIA	0.5589	0.9518	NaN	0.9302
Protein Solubility	4.9391	8.9591	5.8276	1.2310
Residual Phytates	8.3070	1.3284	NaN	0.8058
Residual Tannins	4.6234	0.6451	NaN	0.2499
Residual TIA	0.6766	0.2236	NaN	0.2691

Feature Importance

Both Random Forest and SHAP (SHapley Additive exPlanations) analyses steadily emphasized the importance of Oilcake type and Temperature C as primary drivers for protein solubility prediction.

Table 3.5: Comparison of Random Forest and SHAP Feature Importances for Protein Solubility

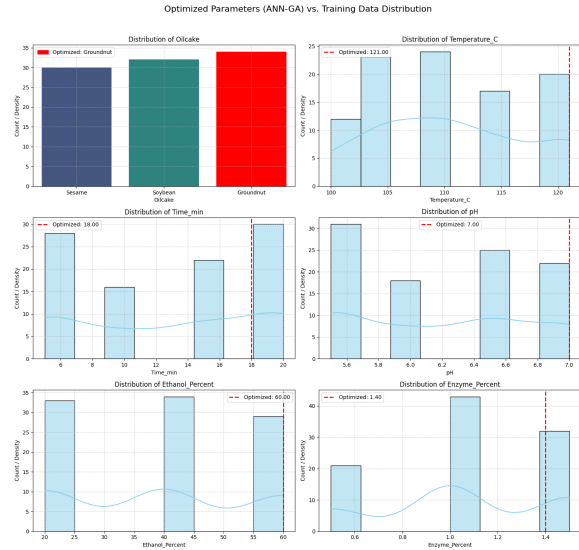
Feature	RF Importance	SHAP Importance
Oilcake	0.3972	1.0806
Temperature C	0.2028	0.4871
Ethanol Percent	0.1259	0.2750
Enzym Percent	0.0995	0.2585
Time (min)	0.1140	0.2022
pH	0.0607	0.0886

Visualizations and Supporting Analysis

Numerous visualizations were produced to provide deeper insights into the optimization process and model performance.

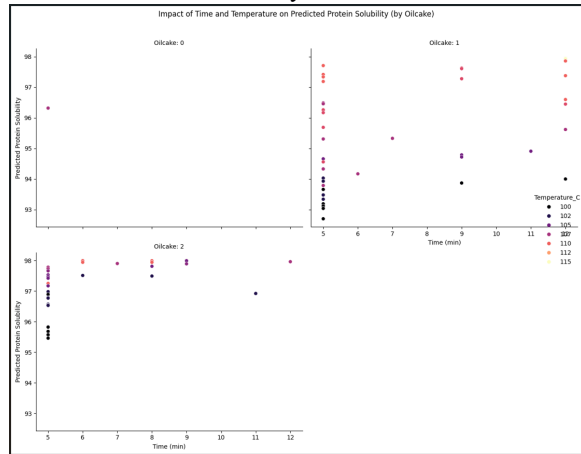
Nutritional Valorization of Oilseed Cakes: A Comparative Study on Proximate Composition, Anti-Nutritional Factors, and AI-Driven Optimization of ANF Reduction Strategies

Figure 1: Optimized Protein Solubility Across Oilcake Types



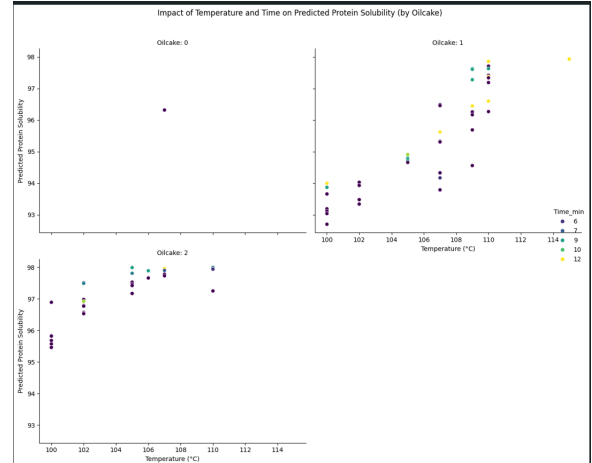
This bar chart displays the mean optimized protein solubility for Groundnut, Sesame, and Soybean oilcakes, highlighting Soybean's superior solubility.

Figure 2: Impact of Temperature and Time on Predicted Protein Solubility



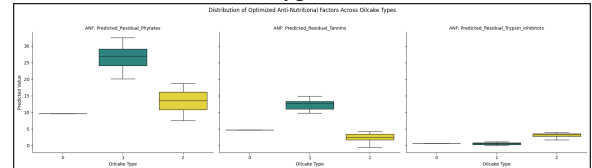
This multi-panel plot illustrates how varying temperature and time affect protein solubility for each oilcake type, with hues indicating time (in minutes).

Figure 3: Impact of Time and Temperature on Predicted Protein Solubility



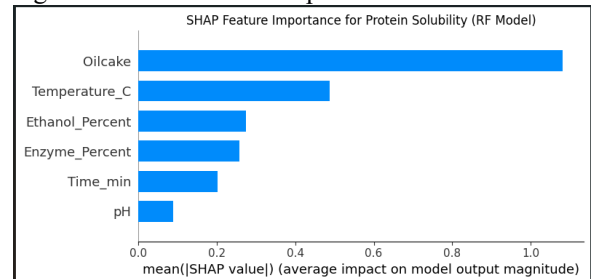
Similar to Figure 2, this plot shows the influence of time and temperature on protein solubility, with hues representing temperature (in Celsius).

Figure 4: Distribution of Optimized Anti-Nutritional Factors Across Oilcake Types



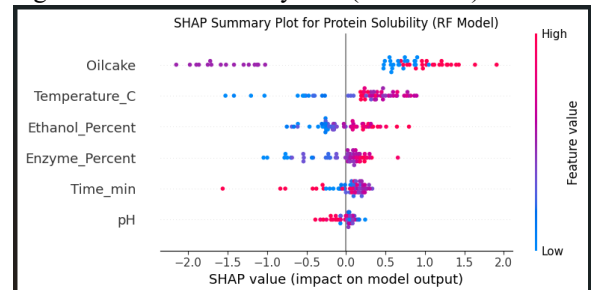
Box plots depicting the distribution of predicted residual phytates, tannins, and trypsin inhibitors across different oilcake types after optimization.

Figure 5: SHAP Feature Importance Bar Plot



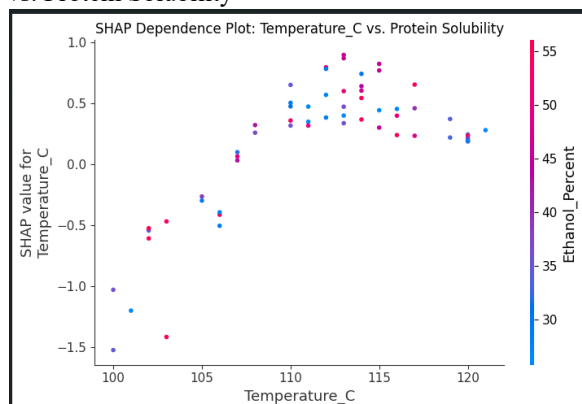
A bar chart illustrating the overall SHAP importance values for each input feature on protein solubility prediction, confirming key drivers.

Figure 6: SHAP Summary Plot (Beeswarm)



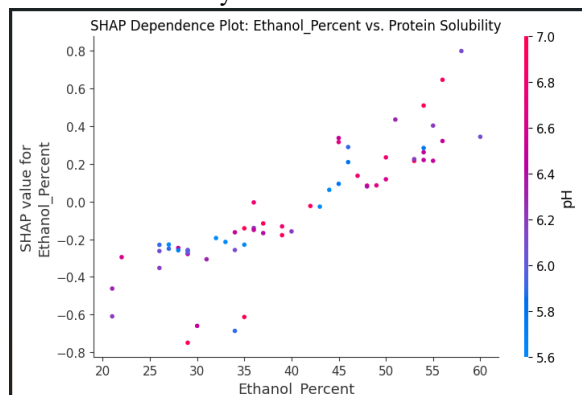
A beeswarm plot showing the distribution of SHAP values for each feature, indicating both importance and the impact (positive/negative) on the model's output.

Figure 7: SHAP Dependence Plot: Temperature C vs. Protein Solubility



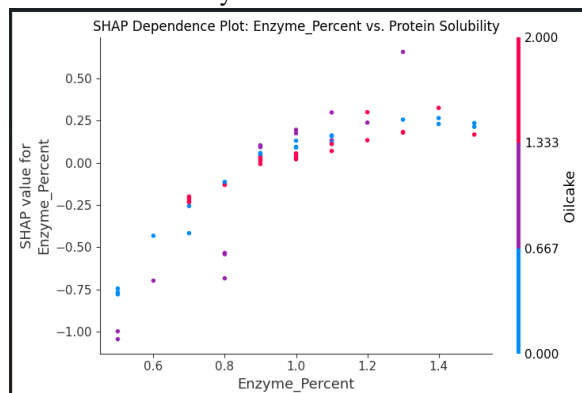
This plot shows how the SHAP values for Temperature C change with its actual values, revealing its interaction effect on protein solubility.

Figure 8: SHAP Dependence Plot: Ethanol Percent vs. Protein Solubility



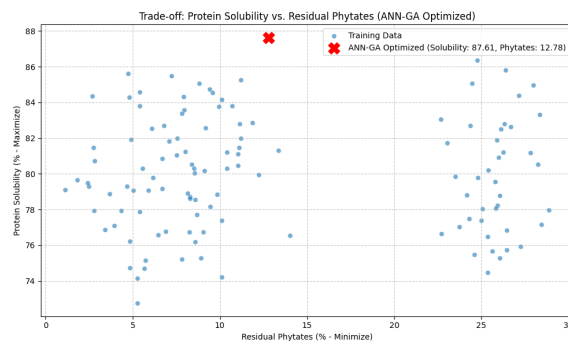
This plot illustrates the marginal effect of Ethanol Percent on protein solubility through its SHAP values.

Figure 9: SHAP Dependence Plot: Enzyme Percent vs. Protein Solubility



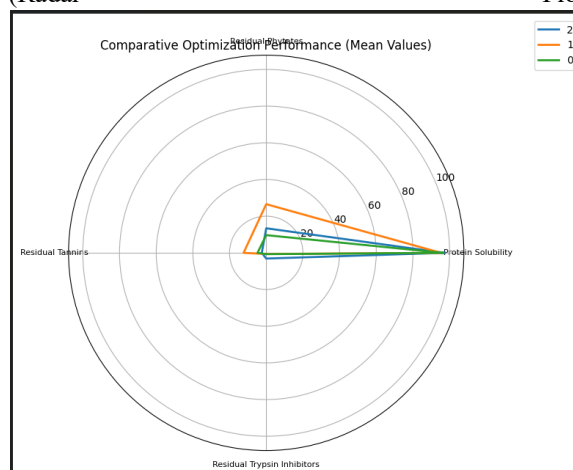
This plot visualizes the influence of Enzyme Percent on protein solubility, as explained by SHAP values.

Figure 10: Trade-Off Between Protein Solubility and Residual Phytates



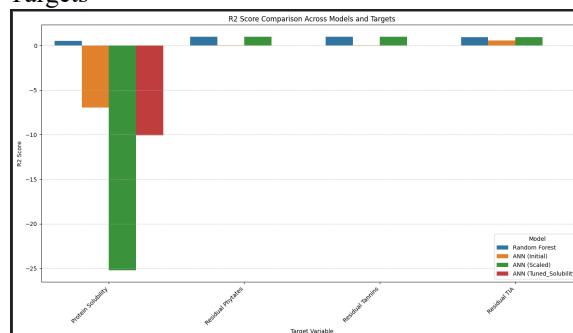
A scatter plot illustrating the inverse relationship between predicted protein solubility and residual phytates, highlighting the trade-off in multi-objective optimization.

Figure 11: Comparative Optimization Performance (Radar Plot



A radar chart providing a holistic view of the average optimized performance across key objectives (protein solubility and ANFs) for each oilcake type.

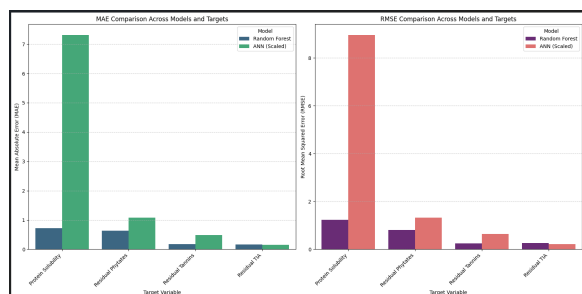
Figure 12: R2 Score Comparison Across Models and Targets



A bar chart comparing the R2 scores of different models (Random Forest, initial ANN, scaled ANN, tuned ANN) across various target variables.

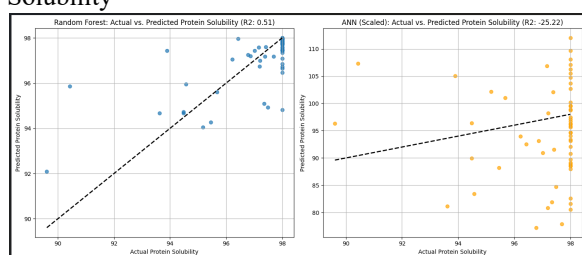
Figure 13: MAE and RMSE Comparison Across Models and Targets

Nutritional Valorization of Oilseed Cakes: A Comparative Study on Proximate Composition, Anti-Nutritional Factors, and AI-Driven Optimization of ANF Reduction Strategies



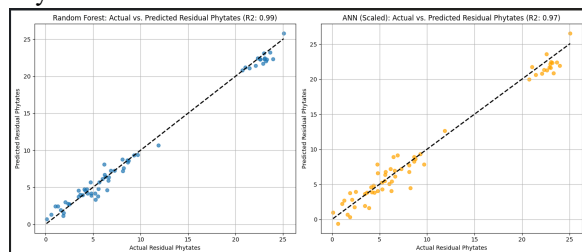
Bar charts comparing the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the Random Forest and scaled ANN models for all target variables.

Figure 14: Actual vs. Predicted Plots for Protein Solubility



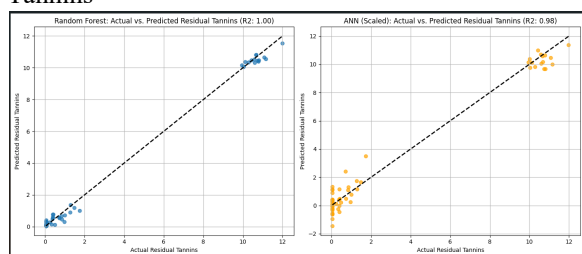
Scatter plots showing the actual vs. predicted values for protein solubility from both Random Forest and ANN models, with R2 scores indicated.

Figure 15: Actual vs. Predicted Plots for Residual Phytates



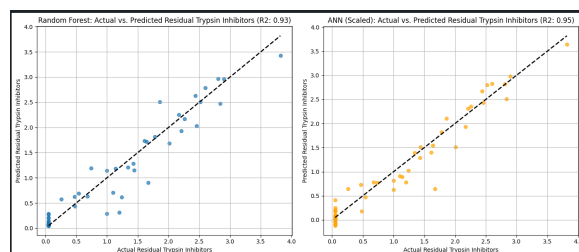
Scatter plots showing the actual vs. predicted values for residual phytates from both Random Forest and ANN models.

Figure 16: Actual vs. Predicted Plots for Residual Tannins



Scatter plots showing the actual vs. predicted values for residual tannins from both Random Forest and ANN models.

Figure 17: Actual vs. Predicted Plots for Residual Trypsin Inhibitors



Scatter plots showing the actual vs. predicted values for residual trypsin inhibitors from both Random Forest and ANN models.

Conclusion

The worldwide demand for plant-based proteins has shifted focus toward agricultural by-products like oilcakes. They are utilized as feed additives because they contain high protein. In food industry, they can be incorporated into fortified flour mixtures, breakfast cereals, snacks, and beverages to enhance protein, fibre, and micronutrient levels. This assists in alleviating malnutrition and enhancing public health benefits. They can also find application in the manufacture of protein isolates and concentrates for food.

Although they are nutritionally rich, the presence of anti-nutritional factors (ANF) limits the bio availability of the nutrients by affecting their digestibility and solubility. Anti-nutritional compounds normally found in oilcakes are trypsin inhibitors, phytates, tannins, along with others like lipoxigenase, oligosaccharides, lectins, and saponins. For efficient utilization of oilcakes in the food industry, it is necessary to remove these ANFs without affecting the protein compounds in them. Although there are several methods like, thermal processing, chemical treatment, enzymatic hydrolysis, soaking, fermentation, roasting, boiling to remove ANFs, it is important to find which method at what processing parameter removes them the most efficiently.

This is where Artificial intelligence (AI) driven optimization of the ANF reduction strategy is used. The optimization was performed using a Genetic Algorithm (GA) driven by Random Forest and Artificial Neural Networks (ANN) models. The process was aimed at enhancing protein solubility and minimizing anti-nutritional factors (ANFs) in Groundnut, Sesame, and Soybean oilcakes. The multi-objective optimization approach effectively identified and validated optimal processing parameters for Groundnut, Sesame, and Soybean oilcakes. This comprehensive study provides actionable insights for improving the nutritional quality of these oilcakes by increasing protein solubility while effectively mitigating anti-nutritional factors. The mix of predictive modeling, genetic algorithms, and thorough interpretability analyses

offers a strong framework for future process development in the food and feed industries. The incorporation of AI into food technology delivers a superior pathway for the valorization of oilseed cakes. This study successfully identified processing interventions that neutralize anti-nutritional factors without compromising the substrate's value for protein hydrolysate production. This computational approach reduces experimental overhead and provides a scalable model for the functional food industry.

Conventional optimization of ANF reduction often depends on linear statistical models which may miss to determine the complex, non-linear thermal degradation kinetics of these compounds. This research incorporates Food Industry 4.0 principles by using AI-based modelling to capture optimum processing parameters.

References:

Anyiam, P. N., Phongthai, S., Sai-Ut, S., Kingwascharapong, P., Jung, Y. H., Zhang, W., & Rawdkuen, S. (2025). Nutritional components and digestibility profiles of some potential plant-based protein sources. *Foods*, 14(10), 1769.

Arsov, A., Tsigoriyna, L., Batovska, D., Armenova, N., Mu, W., Zhang, W., ... & Petrova, P. (2024). Bacterial degradation of antinutrients in foods: The genomic insight. *Foods*, 13(15), 2408.

Baisch, J. S., Grohs, M., Ferreira, P. A. A., Ugalde, G. A., Tres, M. V., & Zabot, G. L. (2024). Protein and oil contents, micro-and macronutrients, and other quality indicators of soybean cultivated in lowland fields. *Foods*, 13(23), 3719.

Bamgbose, A. M. (2013). Chemical analyses of some oilseed-cakes.

Behera, S., Indumathi, K., Mahadevamma, S., & Sudha, M. L. (2013). Oil cakes—a by-product of agriculture industry as a fortificant in bakery products. *International journal of food sciences and nutrition*, 64(7), 806-814.

Das, R., & Bhattacharjee, C. (2015). Processing sesame seeds and bioactive fractions. *Processing and impact on active components in food*, 385-394.

Embaby, H. E. S. (2010). Effect of heat treatments on certain antinutrients and in vitro protein digestibility of peanut and sesame seeds. *Food science and technology research*, 17(1), 31-38.

Gamba, M., Nagar, S., Kumari, K., Tanwar, E., Goyal, S., Kumar, V., & Singh, B. (2024). Cyclic extraction of phosphate from soybean meal using immobilized *Aspergillus oryzae* SBS50 phytase. *Bioprocess and Biosystems Engineering*, 47(1), 39-55.

Gilani, G. S., Cockell, K. A., & Sepehr, E. (2005). Effects of antinutritional factors on protein digestibility and amino acid availability in

foods. *Journal of AOAC international*, 88(3), 967-987.

Katoch, O. R. (2024). Tackling child malnutrition and food security: assessing progress, challenges, and policies in achieving SDG 2 in India. *Nutrition & Food Science*, 54(2), 349-365.

Kaur, M., Singh, B., & Kaur, A. (2022). Influence of different oilseed cake incorporation on batter rheology, proximate composition, texture, antioxidant and sensory properties of wheat flour muffins. *International Journal of Food Science and Technology*, 57(11), 7107-7115.

Khattab, R. Y., Arntfield, S. D., & Nyachoti, C. M. (2009). Nutritional quality of legume seeds as affected by some physical treatments, Part 1: Protein quality evaluation. *LWT-food Science and Technology*, 42(6), 1107-1112.

Latif, K. A., Alam, M. T., Sayeed, M. A., Hussain, M. A., Sultana, S., & Hossain, M. A. (2008). Comparative study on the effects of low cost oil seed cakes and fish meal as dietary protein sources for Labeo rohita (Hamilton) fingerling. *University Journal of Zoology, Rajshahi University*, 27, 25-30.

Lee, J., & Nyachoti, C. M. (2021). Heat processing increased the digestibility of phosphorus in soybean expeller, canola meal, and canola expeller fed to growing pigs. *Journal of Animal Science*, 99(10), skab276.

Li, X., Hua, Y., Chen, Y., Kong, X., Zhang, C., & Yu, X. (2017). An advance for removing antinutritional protease inhibitors: Soybean whey purification of Bowman-Birk chymotrypsin inhibitor by combination of two oppositely charged polysaccharides. *Carbohydrate Polymers*, 164, 349-357.

Manjit, Yadav, A., Aggarwal, N. K., Kumar, K., & Kumar, A. (2008). Tannase production by *Aspergillus fumigatus* MA under solid-state fermentation. *World Journal of Microbiology and Biotechnology*, 24(12), 3023-3030.

Mugendi, J. B., Njagi, E. M., Kuria, E. N., Mwasaru, M. A., Mureithi, I. J. G., & Apostolides, Z. (2010, September). Effects of processing *Mucuna bean* (*Mucuna pruriens* L.) on protein quality and anti-nutrients content. In *Conference on International Research on Food Security, Natural Resource Management and Rural Development* (Vol. 9, pp. 14-16).

Nair, R., Ghakker, N., & Sharma, A. (2015). spectrophotometric estimation of tannins in raw and processed form (Paan Masala) of Areca nut. *Int J Educ Sci Res Rev*, 1, 51-56.

- Nouska, C., Deligeorgaki, M., Kyrkou, C., Michaelidou, A. M., Moschakis, T., Biliaderis, C. G., & Lazaridou, A. (2024). Structural and physicochemical properties of sesame cake protein isolates obtained by different extraction methods. *Food Hydrocolloids*, 151, 109757.
- Ozdal, T., Capanoglu, E., & Altay, F. (2013). A review on protein–phenolic interactions and associated changes. *Food Research International*, 51(2), 954-970.
- Polshettiwar, S. A., Ganjiwale, R. O., Wadher, S. J., & Yeole, P. G. (2007). Spectrophotometric estimation of total tannins in some ayurvedic eye drops. *Indian Journal of Pharmaceutical Sciences*, 69(4).
- Palange, N. J., Obua, T., Sserumaga, J. P., Wembabazi, E., Ochwo-Ssemakula, M., Adjei, E. A., ... & Tukamuhabwa, P. (2025). Unraveling the genetic architecture of anti-nutritional factors in soybean (*Glycine max.*) for nutritional enhancement. *Scientific Reports*, 15(1), 42787.
- Popova, A., & Mihaylova, D. (2019). Antinutrients in plant-based foods: A review. *The Open Biotechnology Journal*, 13(1).
- Popović, S., Hromiš, N., Šuput, D., Bulut, S., Romanić, R., & Lazić, V. (2020). Valorization of by-products from the production of pressed edible oils to produce biopolymer films. In *Cold pressed oils* (pp. 15-30). Academic Press.
- Rakita, S., Kokić, B., Manoni, M., Mazzoleni, S., Lin, P., Luciano, A., ... & Pinotti, L. (2023). Cold-pressed oilseed cakes as alternative and sustainable feed ingredients: A review. *Foods*, 12(3), 432.
- Ramachandran, S., Singh, S. K., Larroche, C., Soccol, C. R., & Pandey, A. (2007). Oil cakes and their biotechnological applications—A review. *Bioresource technology*, 98(10), 2000-2009.
- Salim, R., Nehvi, I. B., Mir, R. A., Tyagi, A., Ali, S., & Bhat, O. M. (2023). A review on anti-nutritional factors: Unraveling the natural gateways to human health. *Frontiers in Nutrition*, 10, 1215873.
- Sessa, D. J., & Nelsen, T. C. (1991). Chemical inactivation of soybean protease inhibitors. *Journal of the American Oil Chemists' Society*, 68(7), 463-470.
- Singh, R., Langyan, S., Sangwan, S., Rohtagi, B., Khandelwal, A., & Shrivastava, M. (2022). Protein for human consumption from oilseed cakes: a review. *Frontiers in Sustainable Food Systems*, 6, 856401.
- Sunil, L., Appaiah, P., Prasanth Kumar, P. K., & Gopala Krishna, A. G. (2015). Preparation of food supplements from oilseed cakes. *Journal of food science and technology*, 52(5), 2998-3005.
- Vichare, S. A., & Morya, S. (2024). Exploring waste utilization potential: nutritional, functional and medicinal properties of oilseed cakes. *Frontiers in Food Science and Technology*, 4, 1441029.
- Zhang, M., Wang, O., Cai, S., Zhao, L., & Zhao, L. (2023). Composition, functional properties, health benefits and applications of oilseed proteins: A systematic review. *Food Research International*, 171, 113061.
- Zhang, Q., Tong, X., Sui, X., Wang, Z., Qi, B., Li, Y., & Jiang, L. (2018). Antioxidant activity and protective effects of Alcalase-hydrolyzed soybean hydrolysate in human intestinal epithelial Caco-2 cells. *Food Research International*, 111, 256-264.