Comparison of Decision Table Method and Random Forest in Providing Funds for Underprivileged Students at Darul Makmur Al-Hafiz Islamic Boarding School, Southeast Aceh

Aidul Safii[1], Zulham Sitorus[2]

Abstract

Offering monetary support to disadvantaged students exemplifies the pesantren's commitment to fostering educational continuity for those encountering financial hardships. Nonetheless, the manual approach to identifying aid beneficiaries can lead to complications, including bias in decision-making, variability in evaluations, and difficulties in concurrently handling multiple criteria. This research is to evaluate the relative efficacy of the Decision Table and Random Forest techniques in determining the eligibility for financial support for disadvantaged students at Pesantren Darul Makmur Al-Hafiz Aceh Tenggara. The research data was derived from student information encompassing many parameters linked to social and economic circumstances, including parental income, number of dependents, parental occupation, housing situation, and additional relevant factors. The research methodology encompasses data collection, data sanitisation and conversion, the creation of classification models employing the Decision Table and Random Forest techniques, and the assessment of the efficacy of these approaches. The models are assessed thru a confusion matrix incorporating metrics like accuracy, precision, recall, and F1 score. The ensuing assessment outcomes are examined to identify the most efficient approach for categorising pupils who qualify and do not qualify for aid. This research aims to yield a more precise, impartial, and uniform categorisation technique, thereby establishing a basis for the creation of a decision support system to allocate funding to disadvantaged students at Pesantren Darul Makmur Al-Hafiz Aceh Tenggara.

***Keywords**:* Decision Table; Random Forest; Classification, Students, Financial Aid, DSS

Aidul Safii[1]
[1]Master's Study Program in Information Technology, Faculty of Postgraduate Studies, Universitas Pembangunan
e-mail: : rahmayunisimanullang2009@gmail.com[1]

Zulham Sitorus[2]
[2,3] Master of Management, Universitas Pembangunan Panca Budi, Indonesia
e-mail zulhamsitorus@dosen.pancabudi.ac.id[2]
3rd International Cofference on the Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Navigating The Future: Business and Social Paradigms in a Transformative Era.
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

Education serves as a vital instrument in enhancing the calibre of human resources and facilitating opportunities for individuals to attain an improved standard of living. In the context of Islamic education in Indonesia, pesantren are vital as they offer religious instruction while simultaneously serving social purposes and fostering communal empowerment. This stance is supported by the Republic of Indonesia Law Number 18 of 2019 about Pesantren, which underscores that pesantren fulfilll the roles of education, religious instruction, and community development (Republic of Indonesia, 2019). Consequently, the presence of pesantren is not solely aimed at enhancing the religious competencies of pupils but also entails a social obligation to facilitate educational opportunities for individuals from many economic strata.

Economic challenges continue to be a significant element influencing the community's capacity to fulfilll educational requirements. [1] indicated that the percentage of impoverished individuals in Aceh Province in March 2025 was 12.33%, with rural areas seeing a rate of 14.44%. This scenario suggests that the challenge of financial constraints continues to be significant, particularly for communities dependent on household earnings to fulfilll their educational requirements.

Disparities in familial financial circumstances are also evident within the realm of pesantren schooling. Certain students originate from secure financial environments, whereas others hail from households who find it challenging to support their educational and personal requirements throughout their time at the pesantren. Consequently, the distribution of financial resources for underprivileged students is an essential means of assistance to guarantee the persistence of their study. This distribution must be executed meticulously to ensure that the available resources are allocated to students who genuinely fulfilll the criteria for help.

The precision in identifying beneficiaries of help is a crucial factor, as these determinations frequently encompass multiple factors. Evaluating an individual's financial status cannot rely on a single factor; it necessitates the simultaneous consideration of multiple socio-economic attributes. In their study on categorising social assistance beneficiaries, [2] employed multiple factors, including income, family size, and living circumstances, as metrics to assess the eligibility of participants. The research suggests that socio-economic information can be utilised as a foundation for categorisation to facilitate more organised decision-making.

Complications emerge when the pool of prospective recipients expands, with each individual subject to distinct conditions influenced by diverse parameters. The assessment procedure that depends entirely on manual examination may need more time and could face challenges in maintaining consistent application of standards. Two students in comparable situations ought to be granted consistent outcomes if the same standards are applied. Consequently, a technique is required that can assess data according to patterns identified in historical records, enabling the feasibility review process to be conducted in a more systematic and quantifiable way.

One approach to tackle the problem is using data mining, particularly employing categorisation tools. Classification is utilised to create a model grounded in classed or labelled data, enabling the model to ascertain the category of data according to its characteristics. This research delineates class categories as those qualified and unqualified for financial assistance, with attributes tailored to the standards or information pertaining to pupils at the Darul Makmur Al-Hafiz Islamic Boarding School in Southeast Aceh.

The utilisation of categorisation methods to identify beneficiaries of aid has been conducted in numerous prior investigations. [2] employed the Random Forest algorithm to classify the eligibility of social aid recipients utilising 1,100 data points with attributes such as income, family size, and living conditions. The research results demonstrated a 97% accuracy rate, suggesting that Random Forest possesses substantial potential for classifying eligibility for help clients. [3] executed a study that juxtaposed the Decision Tree and Random Forest algorithms for categorising recipients of Non-Cash Food Assistance (BPNT) in Slangit Village. The study results demonstrate that Decision Tree and Random Forest yield the same accuracy rate of 97.86% on the examined dataset. The findings are noteworthy as they suggest that an algorithm's efficacy cannot be evaluated exclusively on its theoretical attributes, but must also be examined in relation to the data qualities under investigation. An algorithm demonstrating optimal efficacy on a specific dataset may not exhibit comparable performance when utilised on a dataset with differing attributes [3].

The assessment of the two techniques should be performed with impartial evaluation criteria. The assessment of a classification model should not depend exclusively on the total count of correct predictions, but must also take into account the model's proficiency in recognising each category. [4] elucidate that classification evaluation can be conducted utilising various metrics derived from the confusion matrix. Consequently, this research can evaluate the juxtaposition of Decision Table and Random Forest utilising metrics such as accuracy, precision, recall, and F1-score, ensuring that the optimal approach is not ascertained thru a solitary performance measure. This study aims to develop a model that demonstrates optimal performance according to testing outcomes and furnish a data-driven basis to enhance the selection process for funding beneficiaries among disadvantaged students in a more objective, consistent, transparent, and focused way.

Literature Review

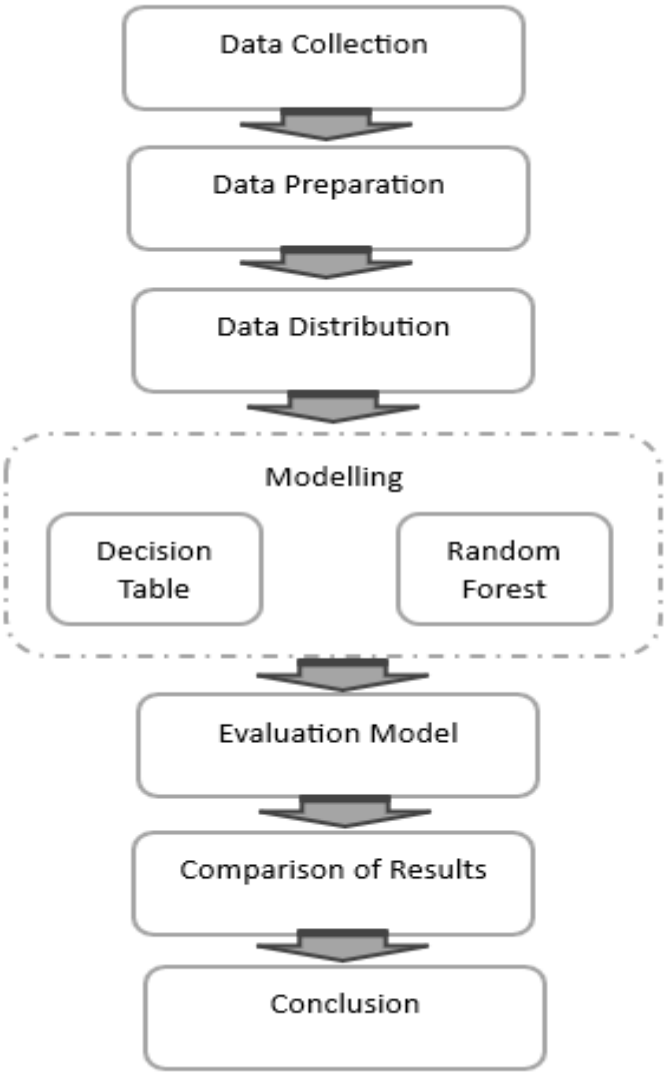
The research on assessing the qualifications of aid beneficiaries has advanced concurrently with the growing application of data mining and machine learning in the decision-making framework. The data-centric methodology facilitates the analysis of historical data regarding prospective recipients to discern particular patterns, which are subsequently utilised in the classification procedure. [5] elucidate that the Knowledge Discovery in Databases (KDD) process entails the identification of patterns within data that are valid, novel, potentially advantageous, and comprehensible. Data mining serves a vital function as a step for uncovering patterns or insights within the dataset.

Data Mining is a data processing approach that employs statistical methods, mathematics, artificial intelligence, and machine learning to extract and discern information from a dataset [6]. An additional characterisation of Data Mining is the exploration of information and analysis of patterns thru certain methodologies or approaches [7]. The primary aim of the classification model is to forecast the value of an unidentified target variable by leveraging data from known variables [8]. Evaluation of data thru a comparative analysis of the efficacy of two data mining algorithms [9].

The intricacy of this selection procedure, which necessitates the assessment of several factors like qualifications, demands a strong Decision Support System (DSS) [10]. Examination and discourse concerning data mining to forecast the volume of data from each Decision to achieve outcomes [11].

Research Methodology

The approach of this study initiates with the stage of gathering information from students concerning the procedure for identifying beneficiaries of funds for the underprivileged. The utilised data encompasses the socio-economic details of the pupils alongside the eligibility criteria for fund beneficiaries derived from historical records, namely the classifications of eligible and ineligible. The subsequent phase encompasses data preparation, which entails cleansing, addressing deficient data, eliminating duplicates, and converting data to satisfy modelling requirements. The organised data is further partitioned into 70% designated for training and 30% allocated for testing. The training dataset is utilised to construct the classification model, whereas the testing dataset is employed to evaluate the model's proficiency in predicting previously unseen data.



Gambar 1. Research Metodology

Results

3.1 Dataset

The dataset employed in this research comprises information on students, encompassing many socio-economic characteristics to assess the eligibility of disadvantaged students for financial support at Pesantren Darul Makmur Al-Hafiz Aceh Tenggara. The dataset comprises 200 students with diverse characteristics, such as parental income, the employment of both parents, the number of dependents in the family, parental status, housing conditions, receipt of social aid, educational fee debts, and residential distance. Every data point is paired with a classification label denoting Eligible or Not Eligible for financial aid. Subsequently, the dataset is partitioned into 70% designated for training and 30% allocated for testing, which will facilitate the creation and assessment of the classification model employing Decision Table and Random Forest techniques. This dataset is employed to evaluate the efficacy of both approaches in determining the eligibility of grant recipients according to the criteria of accuracy, precision, recall, and F1-score. The data from the subsequent study is presented in Table 1.

Table 1. Dataset

ID_Santri	JK	Income Parents	Category Income	Work Father	Job Mother	Amount Dependents	Status House	Condition House
S001	L	2,500,000	1,5-2,5 juta	Farmer	Housewife	7	Own Property	Enough
S002	L	1,500,000	1,5-2,5 juta	Daily Labourer	Small Trader	2	Contract	Enough
S003	P	5,550,000	>3,5 juta	Small Trader	Pegawai Swasta	7	Hitchhiking	Worthy
S004	L	800,000	<1,5 juta	Daily Labourer	Daily Labourer	2	Contract	Less than adequate
S005	L	1,100,000	<1,5 juta	Daily Labourer	Housewife	3	Own Property	Less than adequate
S006	L	4,400,000	>3,5 juta	Driver	Entrepreneur	7	Own Property	Worthy
S007	P	2,350,000	1,5-2,5 juta	Driver	Small Trader	2	Own Property	Enough
S008	P	2,050,000	1,5-2,5 juta	Entrepreneur	Entrepreneur	4	Hitchhiking	Worthy
S009	P	4,900,000	>3,5 juta	Small Trader	Private Employee	4	Hitchhiking	Worthy
S010	P	700,000	<1,5 juta	Driver	Not Working	6	Contract	Enough
S011	L	2,000,000	1,5-2,5 juta	Farmer	Farmer	7	Own Property	Enough
S012	L	2,650,000	2,5-3,5 juta	Private Employee	Housewife	4	Own Property	Enough
S013	L	2,750,000	2,5-3,5 juta	Private Employee	Housewife	7	Hitchhiking	Worthy
S014	L	3,250,000	2,5-3,5 juta	Entrepreneur	Entrepreneur	4	Own Property	Worthy
S015	P	3,450,000	2,5-3,5 juta	Private Employee	Small Trader	2	Own Property	Enough
S016	L	900,000	<1,5 juta	Daily Labourer	Farmer	4	Contract	Enough
S017	L	1,900,000	1,5-2,5 juta	Farmer	Farmer	3	Own Property	Less than adequate
S018	L	4,700,000	>3,5 juta	Entrepreneur	Small Trader	4	Own Property	Worthy
S019	P	4,600,000	>3,5 juta	Entrepreneur	Private Employee	4	Own Property	Worthy
S020	L	1,250,000	<1,5 juta	Not Working	Farmer	4	Hitchhiking	Worthy
S021	P	1,200,000	<1,5 juta	Daily Labourer	Daily Labourer	1	Contract	Less than adequate
S022	P	1,700,000	1,5-2,5 juta	Entrepreneur	Daily Labourer	5	Own Property	Worthy

S023	P	2,550,000	2,5-3,5 juta	Small Trader	Housewife	5	Own Property	Worthy
S024	L	750,000	<1,5 juta	Not Working	Farmer	4	Hitchhiking	Less than adequate

3.2 Preprocessing Data Results

Based on the results of data preprocessing, the dataset used in the research has undergone a stage of completeness and consistency verification before being used in the classification process. From the 200 student data available, there are no missing values or duplicate data, so all data can be retained for the modelling process. In the Decision Table method, several numerical attributes are converted into categorical format to facilitate the formation and interpretation of decision patterns. Age is classified into groups of 13–14 years, 15–17 years, and ≥18 years; the number of dependents is divided into low, medium, and high categories; fee arrears are grouped into none, light, and heavy; while the distance from home is categorised as close, medium, and far. Attributes such as father's profession, mother's profession, parental status, residence status, house condition, social assistance recipient, and income category are used in categorical format, with the target variable being the classes Eligible and Ineligible.

Table 3. Missing Value Handling Results

DESCRIPTION OF DATASET VARIABLES

Variabel	Tipe Data	Role	Explanation	Example Value	Used for Model
ID_Santri	Nominal	Identifier	Unique student code; do not use as a predictor.	SAN001	NO
Gender	Nominal	Prediktor	Gender of the students.	M / P	Yes
Age	Numerik	Prediktor	The age of the students in years.	16	Yes
Parental_Income_Rp	Numerik	Prediktor	Estimated total monthly income of parents.	2.000.000	Yes
Income_Category	Ordinal	Prediktor	Income categories to facilitate the Decision Table.	1,5-2,5 juta	Yes
Father's job	Nominal	Prediktor	The main job of the father/male guardian.	Farmer	Yes
Mother's job	Nominal	Prediktor	The main job of the mother/female guardian.	Housewife	Yes
Number of Dependents	Numerik	Prediktor	The number of family members who are dependents.	5	Yes
Parental_Status	Nominal	Prediktor	The condition of the parents' presence.	Yatim	Yes
House_Status	Nominal	Prediktor	Family residence ownership status.	Hitchhiking	Yes
House Condition	Ordinal	Prediktor	The physical condition of the family house.	Less than adequate	Yes
Social Assistance Recipient	Nominal	Prediktor	Family status receiving social assistance.	Yes	Yes
…	…	…	….	….	….
Subset	Nominal	Data Distribut	Distribution of 70% training and 30% testing in a stratified manner.	Training	No

ion

Conversely, the Random Forest technique preserves numerical features including age, parental income, number of dependents, debt amount, and distance from home in their original numerical format, as this algorithm is capable of directly handling numerical data. Categorical variables are subsequently transformed into numerical representations via binary encoding and one-hot encoding techniques to facilitate model processing. The target variable Kelayakan_Dana is transformed into binary values, assigning Layak = 1 and Tidak Layak = 0. Upon finalising all preparation phases, the dataset is partitioned into 70% designated for training and 30% allocated for testing. An equivalent data distribution is utilised for both algorithms to guarantee that the assessment and comparison of the Decision Table and Random Forest's performance are executed impartially, objectively, and uniformly according to the metrics of accuracy, precision, recall, and F1-score.

DATASET SUMMARY

Indikator	Value
Datos Totales	200
Worthy	104
Not Worthy	96
Training	140
Testing	60

Class	Training	Testing
Worthy	73	31
Not Worthy	67	29
Total	140	60

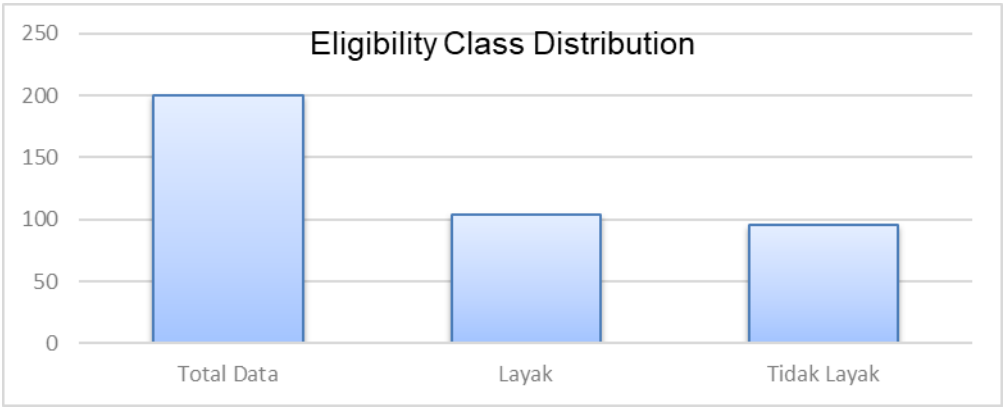


Figure 2. Feasibility Distribution

3.3 Implementasi Algoritma Decision Tabel danRendom Forest

In the Decision Table calculation, the attributes tested as a schema include income category, dependent category, parental status, house status, house condition, social assistance recipient, arrears category, and house distance category. Based on the Leave-One-Out testing on the training data, the best attribute combination is:

Income Category + Liability + House Status

From 140 training data, 117 correct classifications were obtained, thus:

$$\text{Accuracy}_{\text{training}} = \frac{117}{140}\ x\ 100\ \% = 83{,}57\%$$

Income	Category	Liability	House Status	Worthy	Not Worthy	Decision
1,5–2,5 juta		Low	Contract	0	2	Not Worthy
1,5–2,5 juta		Currently	Contract	5	0	Worthy
1,5–2,5 juta		Currently	Own Property	3	7	Not Worthy
1,5–2,5 juta		Tall	Contract	3	0	Worthy
1,5–2,5 juta		Tall	Own Property	8	2	Worthy

In the Random Forest preprocessing results, there are 31 numerical predictor attributes after encoding. For the manual simulation, it is used:

n=140 data training

Amount Tree=50

$mtry = \sqrt{31} \approx 5$

This means that at each node, around 5 attributes are randomly selected to find the best split. Random Forest takes training samples randomly with replacement (bootstrap sampling). In Tree 1, 140 bootstrap samples were obtained with:

Worthy =70

Not Worthy = 70

Then:

$$p_{Layak} = \frac{70}{140} = 0,5$$

$$p_{TidakLayak} = \frac{70}{140} = 0,5$$

Initial node Gini:

$$\text{Gini(root)} = 1-(0,5)2-(0,5)^2$$
$$= 1-0,25-0,25$$

The smaller the Gini value, the more homogeneous the class at that node.

Atribut	Weighted Gini	Gini Gain
Uninhabitable_House_Condition	**0,406215**	**0,093785**
Father's_Job_Driver	0,498698	0,001302
House Condition: Satisfactory	0,498318	0,001682
Status_House_Sharing	0,451066	0,048934
Status_House_Contract	0,489796	0,010204

Metode	Accuracy	Precision	Recall	F1-Score
Decision Table	78,33%	80,00%	77,42%	78,69%
Random Forest	86,67%	89,66%	83,87%	86,67%

Based on the simulation calculations, Random Forest achieved better performance than Decision Table on all four evaluation indicators. The accuracy difference is: 86.67%−78.33%=8.34%.

Thus, in the synthetic dataset of this study, Random Forest provides better classification performance, while Decision Table has the advantage of simpler decision rules that are easier to explain to users.

Conclusion

Based on the processing and testing of the data, both the Decision Table and Random Forest methods can be used to classify the eligibility for funding for underprivileged students at Pesantren Darul Makmur Al-Hafiz Aceh Tenggara. The Decision Table method produced an accuracy of 78.33%, precision of 80.00%, recall of 77.42%, and an F1-score of 78.69%. Meanwhile, the Random Forest method achieved an accuracy of 86.67%, precision of 89.66%, recall of 83.87%, and an F1-score of 86.67%. These results indicate that both methods have fairly good classification capabilities, but Random Forest outperformed in all evaluation metrics used.

Thus, it can be concluded that Random Forest is a superior method compared to Decision Table in classifying students who are eligible and ineligible to receive funds in this research dataset. The difference in accuracy of 8.34% indicates that the ensemble approach in Random Forest is capable of producing more accurate predictions compared to the simpler classification rules in Decision Table. However, Decision Table still has the advantage in terms of interpretation because the decision rules generated are easier to understand and trace. Therefore, if the main priority of the research is to achieve the best classification performance, Random Forest is more recommended, while Decision Table can be used if ease of interpretation and transparency of rules are the main considerations in the decision-making process.

References

[1] S. Daerah, “Provinsi aceh,” 2025.

[2] S. Assistance, R. In, and I. Systems, “Implementation of Random Forest Algorithm for Classification of Eligibility For,” vol. 8, no. 2, pp. 275–286, 2024.

[3] K. Cirebon, K. Klangenan, D. Sosial, K. Cirebon, and D. Tree, “PERBANDINGAN TINGKAT AKURASI ALGORITMA DECISION TREE DAN RANDOM FOREST DALAM MENGKLASIFIKASI PENERIMA BANTUAN SOSIAL BPNT DI DESA SLANGIT I gini (x) = ? x p (x | y)(1 ? p (x | y)),” vol. 8, no. 1, pp. 127–132, 2024.

[4] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” vol. 45, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.

[5] I. Sabri and H. Bahia, “A Data Mining Model by Using ANN for Predicting Real Estate Market : Comparative Study,” vol. 2013, no. October, pp. 162–169, 2013.

[6] J. T. Santoso *et al.*, “Comparison of Classification Data Mining C4 . 5 and Naïve Bayes Algorithms of EDM Dataset,” vol. 10, no. 4, pp. 1738–1744, 2021, doi: 10.18421/TEM104.

[7] S. Kasus, P. Pt, G. Gunadi, and D. I. Sensuse, “PENERAPAN METODE DATA MINING MARKET BASKET ANALYSIS TERHADAP DATA PENJUALAN PRODUK BUKU DENGAN MENGGUNAKAN ALGORITMA APRIORI DAN FREQUENT PATTERN GROWTH (FP-GROWTH) :,” vol. 4, no. 1, 2012.

[8] I. Kegiatan, “Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan

Algoritma C4.5," vol. 9, no. 4, 2022.
[9] K. Neighbors, "Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin Di Kabupaten Labuhanbatu Menggunakan Random Forest Dan," vol. 6, no. 2, pp. 23–36, 2025, doi: 10.47065/bit.v5i2.1783.
[10] Z. Sitorus, A. Karim, A. H. Nasyuha, and H. Aly, "Implementation of MOORA and MOORSA Methods in Supporting Computer Lecturer Selection Decisions," pp. 554–566, 2024.
[11] N. Febrina, H. Br, Z. Sitorus, and R. F. Wijaya, "Data Mining Analysis in Predicting the Number of New Students at IT & B Indonesia Using Linear Regression Method," vol. 3, no. 1, pp. 1–6, 2024.