Indonesian Journal of Data and Science

Volume 6 Issue 1
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i1.240

Research Article

Evaluating Machine Learning Approaches: A Comparative Study of Random Forest and Neural Networks in Grade Classification

Subhita Sivakumar [1,*], Sivakumar Venkataraman [2]

[1] New Era Collega of Arts, 36158 Gaborone, Bostwana, ssivakumar@neweracollega.ac.bw
[2] Botho University, Gaborone, Botswana, sivakumar.venkataraman@bothouniversity.ac.bw

Correspondence should be addressed to Subhita Sivakumar; ssivakumar@neweracollega.ac.bw

Received 22 December 2024; Accepted 20 March 2025; Published 31 March 2025

Abstract:

Introduction: Accurate grade classification in education is essential for early intervention and performance assessment. This study presents a comparative analysis of Random Forest and Neural Networks in classifying student grades using a dataset of 2,392 high school students. The aim is to evaluate both models' predictive performance and interpretability in an educational data mining context. **Methods**: The dataset, containing academic and demographic features, was pre-processed by handling missing values, encoding categorical variables, and scaling numerical features. Grades were categorized into five classes: A, B, C, D, and F. Both models were implemented using Python and evaluated with metrics including accuracy, precision, recall, and F1-score. Hyperparameter tuning was performed via Grid Search with cross-validation to optimize performance. **Results**: The Random Forest model achieved a baseline accuracy of 70.2%, outperforming Neural Networks at 69.1%. After tuning, Random Forest improved to 71.45% accuracy, while Neural Networks reached 70.49%. Both models demonstrated strong precision and recall in identifying failing students (class F), with F1-scores of 0.90 and 0.89, respectively. However, classification of mid-range grades (A to D) remained challenging due to class overlap. Feature importance analysis highlighted interpretability advantages in the Random Forest model. **Conclusions**: Both models are effective for grade classification, with Random Forest offering slightly better accuracy and interpretability. Neural Networks, while slightly less accurate, capture nonlinear relationships effectively post-tuning. The results suggest that model selection should be guided by context-specific needs, balancing performance with transparency. Future work may include ensemble techniques and expanded feature sets to improve classification robustness.

Keywords: Data Pre-processing, Educational Data Mining, Grade Classification, Hyperparameter Tuning, Model Evaluation, Neural Networks, Random Forest.
Dataset link: https://www.kaggle.com/datasets/rabieelkharoua/students-performance-dataset

1. Introduction

In recent years, classification tasks have become central to a myriad of machine learning applications, ranging from defect detection in integrated circuit packaging to disease diagnosis and ecological mapping. Both neural networks and random forest algorithms have demonstrated distinct strengths in addressing complex classification problems. Neural networks—with architectures such as CNNs, ResNet, deep stacked CapsNet, and vision transformers—have been successfully applied in various domains. For example, [1] employed a CNN on an ESP32-CAM for integrated circuit defect classification, achieving an accuracy of 86.1%, while [2] combined ResNet and advanced image processing techniques to detect weeds in vegetables with accuracies exceeding 95%. Furthermore, studies by [3] and [4] have pushed the envelope in fabric defect classification and retinal disease diagnosis, achieving accuracies of 99.8% and up to 98.1%, respectively, with additional contributions from [5] in radar classification using WiSARD neural networks.

Despite these significant advances, a notable research gap remains in the literature: there is a scarcity of comprehensive comparative studies that evaluate the performance of neural network classifiers against random

forest classifiers across diverse applications. While many studies have focused on specific domains—such as integrated circuit defect classification, agricultural weed detection, or medical diagnosis—the majority have not explored a systematic comparison that addresses predictive performance, computational efficiency, and model interpretability. This gap leaves practitioners without clear guidance on choosing between the high accuracy often provided by neural networks and the robustness and transparency characteristic of random forest methods.

Over the past five years, the state of the art in neural network classification has flourished, as evidenced by the recent work of [1]–[5]. In parallel, random forest classifiers have proven their efficacy in various settings, from diagnosing cataract eye disease with an accuracy of 92% and an F1-score of 92.4% [6], mapping forest types along ecological gradients in Pakistan [7], to enhancements in algorithm performance via hierarchical clustering approaches [8] and precise plant functional type classification using spectral libraries [9]. These advancements underline both the potential and the limitations inherent in each approach, emphasizing the need for a unified evaluation framework.

The core problem addressed by this study is the absence of a comprehensive comparative analysis between neural network-based classifiers and random forest classifiers. Practitioners in fields such as healthcare, agriculture, and manufacturing require robust, evidence-based guidelines to determine the most suitable classification method under specific conditions. The lack of such an integrated comparison often results in the selection of suboptimal models, which may lead to inefficient resource utilization and inadequate interpretability of results.

This research aims to bridge the existing gap by conducting a systematic comparative analysis between neural network and random forest classification methods. The study will evaluate and compare predictive performance using metrics such as accuracy, precision, recall, and F1-score, while also examining training time, computational resource requirements, and scalability across diverse benchmark datasets. Additionally, the research will explore model interpretability through analyses of feature importance and decision-making processes, ultimately providing practical guidelines and a decision-making framework to assist practitioners in selecting the most appropriate classification technique for their specific application needs.

2. Method:

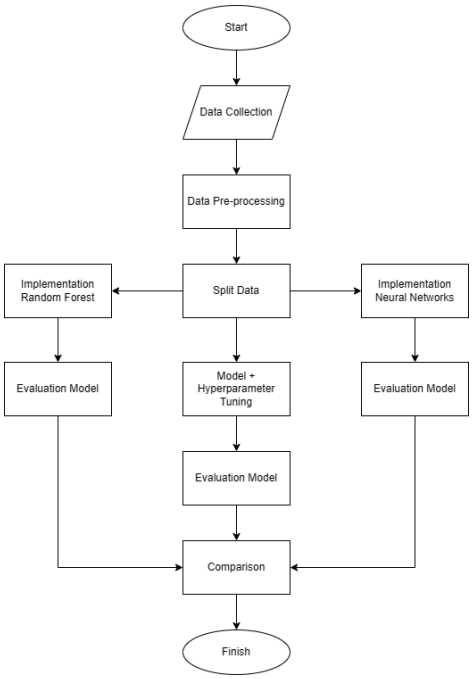

Figure 1. Research Design

Figure 1 illustrates the research design employed in this study, beginning with data collection from relevant sources. Following this, data pre-processing is carried out, which involves handling missing values, encoding categorical variables, and scaling numerical features to ensure the data is suitable for modeling. The dataset is then split into training and testing subsets, after which two main approaches—Random Forest and Neural

Networks—are implemented independently. An initial evaluation of each model's performance is conducted before proceeding to hyperparameter tuning, where key parameters are optimized to enhance predictive accuracy. The tuned models undergo a final evaluation, and their results are compared to identify differences in performance and interpretability, culminating in conclusions and recommendations regarding the most appropriate classification approach for the dataset in question.

Data Collection

In this phase, the dataset is acquired from a comprehensive source containing information on 2,392 high school students. The dataset includes details such as student demographics, study habits, parental involvement, extracurricular activities, and academic performance. The target variable, GradeClass, is derived from the students' GPA and segmented into categories (A, B, C, D, F) based on predefined thresholds. This rich dataset serves as the foundation for both predictive modeling and comparative analysis of classification algorithms.

Data Pre-processing

The pre-processing stage involves several steps to ensure the data is clean and suitable for modeling. Initially, the dataset is examined for missing values and outliers. Missing data are handled using imputation techniques or by removing affected records, depending on the context. Categorical variables such as Gender, Ethnicity, ParentalEducation, and ParentalSupport are transformed using encoding methods (e.g., one-hot encoding or label encoding).

A critical part of pre-processing is data scaling. Scaling is especially important for algorithms like neural networks that are sensitive to the range of input features. The most common technique used is standardization, which transforms each numerical feature to have a mean of zero and a standard deviation of one [10]–[13]. This process is mathematically represented as:

$$x_{scaled} = \frac{x - \mu}{\sigma} \tag{1}$$

Where $\boldsymbol{x}$ is the original pixel value, $\boldsymbol{\mu}$ is the mean, and $\boldsymbol{\sigma}$ is the standard deviation. This transformation ensures that all features contribute equally to the model's learning process, improves the convergence rate during training, and helps avoid issues caused by differing feature scales. It is important to compute these scaling parameters from the training set and apply them consistently to both the training and testing sets to prevent data leakage.

Finally, the dataset is split into training and testing subsets (commonly using a 70:30 ratio) to facilitate unbiased evaluation of the models.

Implementation Algorithm

Two classification algorithms are implemented in this study: Random Forest and Neural Networks. For the Random Forest classifier, an ensemble of decision trees is constructed [14]–[17]. Each tree T_i n the ensemble makes its own prediction $h_i(x)$ for an input x and the final prediction $\hat{y}$ is obtained via majority voting:

$$\hat{y} = model\{h_1(x), h_2(x), \dots, h_N(x)\} \tag{2}$$

This approach leverages the diversity of multiple decision trees to improve robustness and generalization. For the Neural Network classifier, a multi-layer perceptron (MLP) is employed [18]–[22]. The forward propagation in the network is defined by:

$$\begin{aligned} z^{[l]} &= W^{[l]} a^{[l-1]} + b^{[l]} \\ a^{[l]} &= f(z^{[l]}) \end{aligned} \tag{3}$$

Where,

$W^{[l]}$ and $b^{[l]}$ are the weight matrix and bias vector at layer l,

$a^{[l-1]}$ is the activation from previous layer,

$f(\cdot)$ is the activation function (such us ReLU or sigmoid) user at layer l.

The network is trained by minimizing a loss function (typically cross-entropy loss for classification tasks) using

Hyperparameter Tuning

An essential part of the methodology is optimizing model hyperparameters through Grid Search combined with cross-validation [23], [24]. For the Random Forest model, hyperparameters such as the number of trees ($n_{estimator}$), maximum tree depth (max$_depth$) nd the minimum number of samples required to split a node (min$_sample_split$) are tuned. Similarly, for the Neural Network, hyperparameters including the number of hidden layers, the number of neurons per layer, learning rate, and regularization parameters are adjusted [25].

This tuning process can be mathematically described as finding the optimal set of hyperparameters θ^* that minimizes the cross-validation loss:

$$\theta^* = \arg min\ CV\_Loss(\theta) \quad (4)$$

Data Analysis Method

After training, both models are evaluated using standard performance metrics. The evaluation metrics include accuracy, precision, recall, and F1-score. Additionally, confusion matrices are constructed for both models to analyze the distribution of true positives, true negatives, false positives, and false negatives. This analysis provides insights into the types of errors made by each model.

Furthermore, statistical tests may be applied to determine whether the differences in performance metrics are statistically significant. The comprehensive analysis includes not only quantitative performance but also an examination of computational efficiency (e.g., training time and resource usage) and model interpretability. Feature importance analyses, particularly for the Random Forest model, are conducted to understand the influence of each feature on the classification outcomes.

3. Results and Discussion

Results

The experimental evaluation revealed that both the Random Forest and Neural Networks models achieved comparable performance on the grade classification task. Initially, the Random Forest model reached an overall accuracy of approximately 70.2%, while the Neural Networks model attained around 69.1%. Detailed classification metrics indicate that both models perform exceptionally well in identifying the failing grade (class F), with high precision and recall values, whereas the performance for mid-range grades (A, B, C, and D) was more moderate. For instance, **Table 1** shows the classification report for the Random Forest model, where class F achieved a precision of 0.85 and a recall of 0.95, resulting in an F1-score of 0.90. Similarly, **Table 2** presents the classification report for the Neural Networks model, which, while slightly lower overall, still demonstrates strong performance for class F.

Table 1. Classification Report of the Random Forest

	Precision	Recall	F1-score	Support
A	0.38	0.09	0.15	33
B	0.52	0.53	0.52	80
C	0.5	0.55	0.53	121
D	0.55	0.43	0.48	127
F	0.85	0.95	0.9	357
Accuracy			0.7	718
Macro avg	0.56	0.51	0.51	718
Weighted avg	0.68	0.7	0.68	718

Table 2. Classification Report of the Neural Networks

	Precision	Recall	F1-score	Support

A	0.39	0.21	0.27	33
B	0.51	0.53	0.52	80
C	0.5	0.49	0.5	121
D	0.5	0.43	0.46	127
F	0.85	0.94	0.89	357
Accuracy			0.69	718
Macro avg	0.55	0.52	0.53	718
Weighted avg	0.67	0.69	0.68	718

In addition to the classification reports, confusion matrices (referenced in **Figures 2** and **3**) further illustrate the distribution of misclassifications across the grade categories. The majority of the errors occur between adjacent classes, suggesting that the subtle differences in GPA thresholds pose challenges in distinguishing mid-range grades.

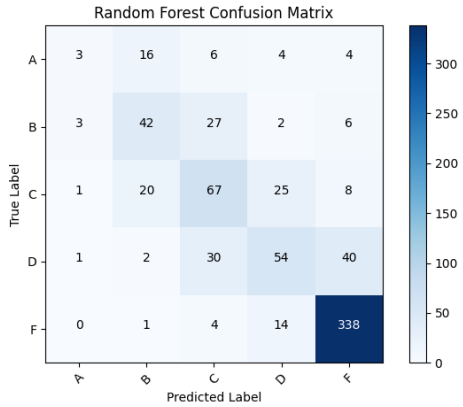

Figure 2. Confusion Matrix of Random forest

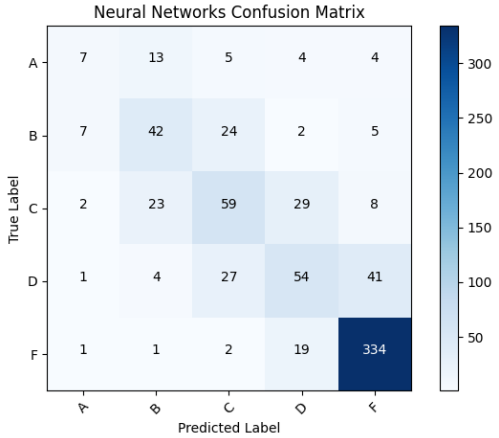

Figure 3. Confusion Matrix of Neural Networks

A visual comparison in **Figure 4** confirms that the baseline accuracies for the two models are very similar, with Random Forest at about 70.2% and Neural Networks at 69.1%. To further enhance model performance, hyperparameter tuning was conducted using Grid Search combined with cross-validation.

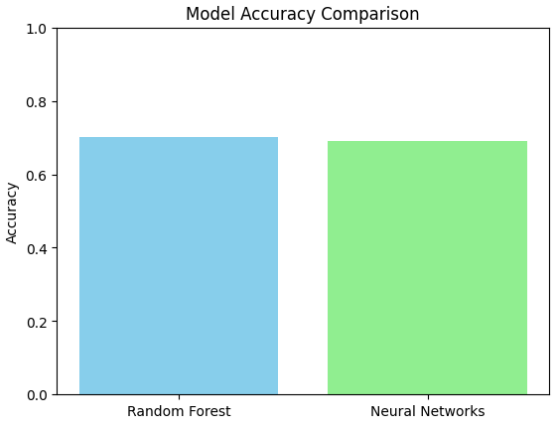

Figure 4. Comparison of Validation Accuracy

As shown in **Table 3**, the Random Forest model's accuracy improved from 0.7019 to 0.7145, while the Neural Networks model's accuracy increased from 0.6908 to 0.7049 after tuning, with the optimal parameters identified for each model.

Table 3. Comparison of Accuracies Before and After Hyperparameter Tuning

Model	Before	After
Random Forest	0.7019	0.7145
Neural Networks	0.6908	0.7049

Discussion

The results indicate that both Random Forest and Neural Networks are effective for the grade classification task, with each model exhibiting distinct strengths and limitations. The Random Forest model not only achieved a slightly higher baseline accuracy but also demonstrated strong performance in classifying failing students, which is critical for early intervention strategies in educational settings. Its inherent interpretability through feature importance analysis adds to its appeal, especially when decision-makers require clear insights into the factors driving classification outcomes.

On the other hand, the Neural Networks model, though initially trailing in performance, showed competitive results following hyperparameter tuning. Its ability to model complex, nonlinear relationships is evident from the improvement in accuracy after fine-tuning key parameters such as network architecture, learning rate, and regularization strength. The observed performance improvements in both models—evidenced by the increase in accuracy after hyperparameter tuning—underscore the importance of optimizing model parameters to extract the best performance from the data.

The analysis also reveals that both models tend to struggle with distinguishing between adjacent grade classes (A, B, C, and D), likely due to the subtle differences in GPA thresholds. However, the robust identification of class F across both models suggests that they are particularly adept at flagging students who are at risk. This observation has significant practical implications, as it supports the use of these models in targeted intervention programs.

In summary, while the Random Forest model offers the benefits of interpretability and slightly higher accuracy, the Neural Networks model provides a viable alternative when complex pattern recognition is required. The choice between these models may ultimately depend on specific operational needs, including the balance between accuracy, interpretability, and computational resources. Future work could explore ensemble approaches or integrate additional feature engineering techniques to further enhance the classification performance across all grade categories.

4. Conclusion

This study conducted a comprehensive comparative analysis of Random Forest and Neural Networks for grade classification using a dataset of 2,392 high school students. The investigation involved meticulous data collection and pre-processing, including scaling for numerical features and encoding for categorical variables, to ensure that

the models could learn effectively from the diverse input data. Both models were implemented and evaluated using standard performance metrics, with particular emphasis on their ability to correctly classify student grades, especially in identifying at-risk individuals.

The experimental results demonstrated that both classifiers achieve comparable accuracy, with the Random Forest model attaining a slightly higher baseline accuracy and improved performance following hyperparameter tuning. The Neural Networks model, while initially trailing, showed significant improvement after fine-tuning key parameters, reflecting its potential to capture complex, nonlinear relationships within the data. Despite their strengths, both models exhibited challenges in differentiating between adjacent grade categories, underscoring the need for further refinement in feature engineering and model optimization.

Overall, the findings suggest that the choice between Random Forest and Neural Networks should be guided by specific operational requirements. Random Forest is preferable in scenarios where model interpretability and ease of deployment are critical, whereas Neural Networks may be more advantageous when the data contains intricate patterns that require advanced nonlinear modeling. Future research could explore hybrid or ensemble methods, as well as additional strategies for feature selection and dimensionality reduction, to further enhance classification performance and support more nuanced decision-making in educational settings.

References:

[1] M. A. Kamaruddin, M. S. Mispan, A. Z. Jidin, H. M. Nasir, and N. I. M. Nor, "Low-cost integrated circuit packaging defect classification system using edge impulse and ESP32CAM," *Int. J. Electr. Comput. Eng.*, vol. 15, no. 1, pp. 156–162, 2025, doi: 10.11591/ijece.v15i1.pp156-162.

[2] H. Jin, K. Han, H. Xia, B. Xu, and X. Jin, "Detection of weeds in vegetables using image classification neural networks and image processing," *Front. Phys.*, vol. 13, Jan. 2025, doi: 10.3389/fphy.2025.1496778.

[3] H. Pooja and S. Soma, "Enhanced Deep Stacked <scp>CapsNet</scp> Ensemble Gazelle Neural Network for multi-level fabric defect classification," *Color. Technol.*, Jan. 2025, doi: 10.1111/cote.12805.

[4] E. S. Cutur and N. G. Inan, "Multi-class Classification of Retinal Eye Diseases from Ophthalmoscopy Images Using Transfer Learning-Based Vision Transformers," *J. Imaging Informatics Med.*, Jan. 2025, doi: 10.1007/s10278-025-01416-7.

[5] M. De Gregorio and M. Giordano, "An experimental evaluation of weightless neural networks for multi-class classification," *Appl. Soft Comput.*, vol. 72, pp. 338–354, Nov. 2018, doi: 10.1016/j.asoc.2018.07.052.

[6] L. Novita, W. Fuadi, and K. Kurniawati, "Cataract Eye Disease Diagnosis Using the Random Forest Method," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 33–41, Jan. 2025, doi: 10.52088/ijesty.v5i2.777.

[7] N. Ahmad and S. G. Ali, "Mapping forest types along ecological gradient in Pakistan," *Environ. Res. Commun.*, vol. 7, no. 3, p. 035023, Mar. 2025, doi: 10.1088/2515-7620/adaf11.

[8] W. Zhuo and A. Ahmad, "HCRF: an improved random forest algorithm based on hierarchical clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, p. 578, Apr. 2025, doi: 10.11591/ijeecs.v38.i1.pp578-586.

[9] A. Mohanta *et al.*, "Harnessing Spectral Libraries From AVIRIS-NG Data for Precise PFT Classification: A Deep Learning Approach," *Plant. Cell Environ.*, Jan. 2025, doi: 10.1111/pce.15393.

[10] M. Sholeh, "Comparison of Z-score, min-max, and no normalization methods using support vector machine algorithm to predict student's timely graduation," *AIP Conference Proceedings*, vol. 3077, no. 1. 2024, doi: 10.1063/5.0202505.

[11] S. Balaji, "Enhancing Diabetic Retinopathy Image Classification using CNN, Resnet, and Googlenet Models with Z-Score Normalization and GLCM Feature Extraction," *Proceedings of the 2nd International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2024*. 2024, doi: 10.1109/IITCEE59897.2024.10467709.

[12] D. Qi, "Improving Unbalanced Security X-Ray Image Classification Using VGG16 and AlexNet with Z-Score Normalization and Augmentation," *Lecture Notes in Electrical Engineering*, vol. 1182. pp. 205–217, 2024, doi: 10.1007/978-981-97-1463-6_14.

[13] D. Geem, "Progression of Pediatric Crohn's Disease Is Associated With Anti–Tumor Necrosis Factor

Timing and Body Mass Index Z-Score Normalization," *Clin. Gastroenterol. Hepatol.*, vol. 22, no. 2, pp. 368–376, 2024, doi: 10.1016/j.cgh.2023.08.042.

[14] A. Faradibah, D. Widyawati, A. U. T. Syahar, and ..., "Comparison Analysis of Random Forest Classifier, Support Vector Machine, and Artificial Neural Network Performance in Multiclass Brain Tumor Classification," *Indones. J. ...*, 2023, doi: 10.56705/ijodas.v4i2.73.

[15] D. Ghunimat, "Prediction of concrete compressive strength with GGBFS and fly ash using multilayer perceptron algorithm, random forest regression and k-nearest neighbor regression," *Asian J. Civ. Eng.*, vol. 24, no. 1, pp. 169–177, 2023, doi: 10.1007/s42107-022-00495-z.

[16] P. Palimkar, R. N. Shaw, and A. Ghosh, "Machine Learning Technique to Prognosis Diabetes Disease: Random Forest Classifier Approach," 2022, pp. 219–244.

[17] M. R. Krause, "Random forest regression for optimizing variable planting rates for corn and soybean using topographical and soil data," *Agron. J.*, vol. 112, no. 6, pp. 5045–5066, 2020, doi: 10.1002/agj2.20442.

[18] M. Bejiga, A. Zeggada, A. Nouffidj, and F. Melgani, "A Convolutional Neural Network Approach for Assisting Avalanche Search and Rescue Operations with UAV Imagery," *Remote Sens.*, vol. 9, no. 2, p. 100, Jan. 2017, doi: 10.3390/rs9020100.

[19] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Math. Comput. Simul.*, vol. 131, pp. 88–100, Jan. 2017, doi: 10.1016/j.matcom.2015.05.010.

[20] D. Gholamiangonabadi, "Deep Neural Networks for Human Activity Recognition with Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection," *IEEE Access*, vol. 8, pp. 133982–133994, 2020, doi: 10.1109/ACCESS.2020.3010715.

[21] P. Henrique Ponte de Lucena and L. Mauro Lima de Campos, "Classification of Obesity Level Using Deep Neural Networks," 2024, pp. 99–107.

[22] A. R. Bhamare, S. Kathargуppe, and J. Silviya Nancy, "Deep Neural Networks for Lie Detection with Attention on Bio-signals," *2020 7th Int. Conf. Soft Comput. Mach. Intell. ISCMI 2020*, no. November 2020, pp. 143–147, 2020, doi: 10.1109/ISCMI51676.2020.9311575.

[23] R. Ghawi and J. Pfeffer, "Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity," *Open Comput. Sci.*, vol. 9, no. 1, pp. 160–180, Jan. 2019, doi: 10.1515/comp-2019-0011.

[24] A. R. Manga, M. A. F. Latief, A. W. M. Gaffar, and ..., "Hyperparameter Tuning of Identity Block Uses an Imbalance Dataset with Hyperband Method," *2024 18th ...*, 2024, doi: 10.1109/IMCOM60618.2024.10418427.

[25] M. Ahsan, M. Mahmud, P. Saha, K. Gupta, and Z. Siddique, "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance," Technologies, vol. 9, no. 3, p. 52, Jul. 2021, doi: 10.3390/technologies9030052.