Received: April 21, 2024. Revised: May 24, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Unleashing SMOTE and Grey Wolf Optimization Approaches to Elevate Neural Networks for Superior Rainfall Prediction

H. Mancy[1, 3]* Amira El Khateeb[2] Hoda A. Ali[3] Kamal El Dahshan[4]

[1]*Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-kharj, Saudi Arabia*
[2] *Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt*
[3] *Department of Mathematics, Faculty of Science (Girls), Al-Azhar University, Cairo, Egypt*
[4]*Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt*
* Corresponding author's Email: h.mancy@psau.edu.sa

Abstract: Precisely predicting rainfall based on extensive meteorological data encompassing various variables poses challenges to achieving high accuracy using traditional machine learning methods. This study uses a comprehensive rainfall dataset derived from weather measurements spanning multiple cities across Australia over the past decade. Its focus lies in optimizing accuracy and minimizing evaluation errors by leveraging the Grey Wolf Optimization algorithm. The primary aim of this algorithm is to feature selection from the dataset, where the fitness function is assessed through machine learning models. These models were individually assessed and also in a hybrid form. The study reveals that the most effective model for rainfall prediction is the SMOTE-Grey Wolf Optimization-Neural Networks (SGWNN) model, showcasing an impressive accuracy of 99.89%. The performance evaluation of these models employed various statistical measures, including Mean Absolute Error, Root Mean Absolute Error, Recall, Precision, Sensitivity, Specificity, and R-squared.

Keywords: Rainfall in Australia, Grey wolf optimization, Machine learning, Neural network, Feature selection.

1. Introduction

Machine learning algorithms play a pivotal role in constructing predictive models or solving complex problems by analyzing extensive datasets. These algorithms, through their ability to discern patterns and relationships within data, generate models that encapsulate the inherent structures. Once developed, these models become powerful tools, efficiently processing information and providing valuable insights or predictions to address various tasks or challenges effectively [1].

Recent focus has been on predicting environmental indicators from datasets, with rainfall as a crucial factor impacting daily life significantly [2]. Its socioeconomic effects include transport disruptions and flood-related infrastructure damage. Climate change exacerbates extreme weather, especially floods, with anticipated catastrophic outcomes [3, 4]. Recent studies show air pollutants spike due to weather variations, affecting respiratory health [5, 6]. Environmental organizations seeking to enhance their performance, attain superior outcomes, and maintain their competitiveness in the ever-changing business landscape of today must priorities optimization. Heuristic optimization techniques, such as the Sunflower Algorithm, provide practical and efficient solutions to difficult optimization issues, particularly in fields where conventional optimization techniques are impractical or inappropriate [7]. A new model using Arctic Sea ice data has improved predictions of Indian Ocean Sea surface temperature, by using deep learning and swarm optimization [8]. Rainfall plays a vital role in agricultural and industrial landscapes, leading to specialized institutes studying these patterns [9]. Predicting rainfall occurrence remains a key focus [10]. Initially, fluid dynamics and thermodynamic

models were used for forecasts. Yet, technological progress has birthed advanced atmospheric models that consider pressure, temperature, and wind, enhancing rainfall understanding for better predictions. These models integrate satellite data, using images to assess clouds, and foresee condensation and precipitation likelihood [11, 12]. This study investigated the effects of Grey Wolf Optimization (GWO) and Synthetic Minority Oversampling Technique (SMOTE) on various machine learning models. Previous studies have investigated the impact of machine learning models without considering whether the dataset is balanced or unbalanced and have not utilized any metaheuristic methods that are used to efficiently solve complex optimization problems where traditional methods may be impractical due to the high dimensionality of the search space or the computational cost of evaluating potential solutions. The objective is to achieve highly accurate early rainfall prediction, specifically in Australia, over the past decade. Previous studies, detailed in Section 2, struggled due to data complexity and the imbalance between rain and no rain statuses. This paper aims to boost prediction accuracy by introducing a novel method: the Grey Wolf Optimization (GWO) algorithm for feature selection, combined with advanced machine learning models like Support Vector Machine (SVM), Random Forest (RF), the eXtreme Gradient Boost technique (XGBoost), and Neural Networks (NNs). It rigorously compares these models with or without GWO on balanced and unbalanced datasets. The collaborative effort culminates in identifying an efficient rainfall prediction method: the innovative GWO-Neural Network model when applied to balanced Synthetic Minority Oversampling Technique (SMOTE) data, showcasing superior performance after comprehensive model comparisons. The fusion of SMOTE with Grey Wolf Optimization within Neural Networks (SGWNN) offers several crucial advantages:

1. **Addressing Imbalanced Data**: SMOTE resolves class imbalance by generating synthetic samples, aiding GWO-optimized neural networks to learn from a more balanced dataset, and ensuring fair representation of minority and majority classes.
2. **Enhanced Model Performance**: GWO fine-tunes neural network parameters while SMOTE balances the dataset, collectively leading to improved predictive accuracy, reduced bias, and increased robustness against imbalanced data.
3. **Improved Generalization**: The combination of SMOTE and GWO results in neural networks that generalize better to unseen data, thanks to reduced bias towards the majority class and a more balanced learning process.
4. **Robustness Against Class Imbalance**: By mitigating the effects of imbalanced data, this approach creates neural networks capable of making more reliable predictions, especially in scenarios where minority class instances are crucial but underrepresented.

This study enhances model performance by introducing the Grey Wolf Optimization (GWO) algorithm for optimal feature selection. Using the Neural Network model on balanced modified data with these selected features achieves an impressive 99.89% accuracy, showcasing a significant advancement in predictive capability.

This paper is organized to achieve its research objectives. Section 2 details the proposed framework architecture. Section 3 presents the methodology of the used model. Section 4 discusses the results of applying the framework. Finally, Section 5 concludes with the findings and suggests future research directions in this domain.

Tabel 1. The Final ML models for rainfall prediction

References	Accuracy on the same dataset	
	Models	**Accuracy**
[13]	Active Algorithm	82%
[13], [14]	Twarit's work	91%
[11]	KNN	83%
[11]	Decision Tree	83%
[11]	Random Forest	83%
[11]	Neural Networks	84%
[15]	XGBoost	90.46%
[15]	Light Gradient Boosting Machine	90.83%
[15]	Random Forest	90.99%
[16]	EK-Stars based LR	87.15%

2. Related work

The following recent studies also used the same dataset that I used for predicting rainfall in Australia at. https://www.kaggle.com/jsphyg/weather-datasetrattle-package#weatherAUS.csv, accessed on March 10, 2022 [28].

2.1 Active learning algorithm (2021 [13])

Active learning algorithms are typically considered semi-supervised rather than purely supervised or unsupervised. Using active learning algorithms with entropy sampling focuses on uncertain instances and targets difficult examples for improved model robustness. Using active learning algorithms with pool-based sampling selects

informative instances from an unlabeled pool and ensures diverse data representation and adapts to data changes. Combining these strategies leads to improved model performance and accuracy, efficient resource utilization by selecting valuable data and Adaptability to evolving data distributions.

- Objective: Prediction of rainfall in Australia using meteorological properties.
- Methodology: Utilization of Active Learning Algorithm with Entropy Sampling and Pool-Based Sampling.
- Classification model Used: Logistics Regression Model.
- Comparison: Active algorithm vs. random sampling; identification of superior sampling method.

2.2 Comparison of classical machine learning techniques (2022 [11])

In this study, each algorithm has its strengths and weaknesses. k-Nearest Neighbors is simple but can be computationally expensive, Random Forest is robust and handles high-dimensional data well, Decision Trees are easy to interpret but prone to overfitting, and Neural Networks are powerful for complex tasks but require careful tuning and sufficient data.

- Objective: Prediction of rainfall in Australia based on meteorological properties.
- Models Compared: the *k*-nearest neighbors' algorithm (*k*-NN), Random Forest, Decision Tree, and Neural Networks.
- Result: Determination of the best-performing model (Neural Networks).

2.3 The eXtreme gradient boost technique vs. random forest (2023 [15])

XGBoost (eXtreme Gradient Boost technique) and Random Forest are both powerful algorithms with their strengths and weaknesses. XGBoost is known for its high performance and scalability, making it suitable for complex tasks and large datasets. On the other hand, Random Forest is more interpretable and can handle high-dimensional data efficiently. The choice between them depends on factors such as the specific task, dataset characteristics, interpretability requirements, and computational resources available.

- Objective: Next-day rainfall status prediction in Australia.
- Comparison: The eXtreme Gradient Boost technique (XGBoost) vs. Random Forest (RF) based on meteorological properties.
- Outcome: Identification of Random Forest as the more accurate predictor.

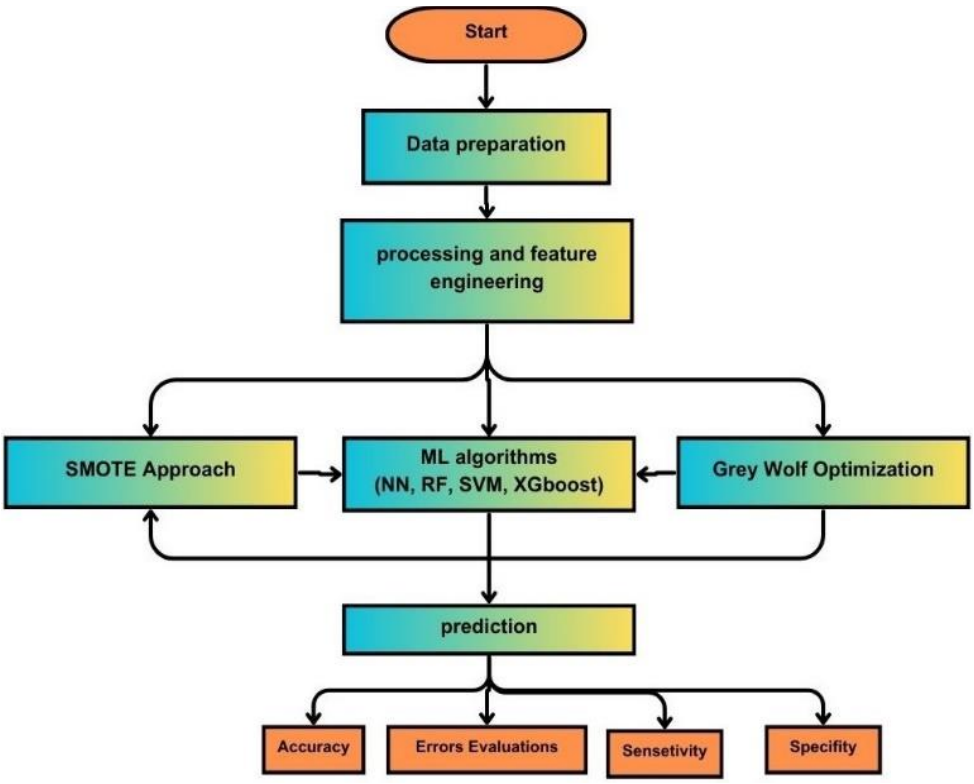

Figure. 1 Proposed Model for Rainfall Prediction

2.4 An ensemble of K-Stars method (2023 [16])

The Ensemble of k-Stars method leverages the strengths of k-NN while addressing its limitations through ensemble learning, leading to more robust and accurate predictions, particularly in scenarios with diverse or noisy data.

- Objective: Next-day rainfall status prediction using ensemble learning.
- Methodology: An Ensemble of K-Stars **(**EK-Stars) method involving ensemble K-Star classifiers.
- Experimentation: Consideration of various scenarios for accurate classification.

2.5 Support vector machine classifier vs. linear regression (2023 [21])

SVM Classifier and Linear Regression are effective machine learning models with distinct characteristics. SVM is suitable for classification tasks with complex decision boundaries and moderate-sized datasets, while Linear Regression is ideal for regression tasks with linear relationships and interpretable models. The choice between them depends on the specific task requirements, data characteristics, interpretability needs, and computational resources available.

- Objective: Prediction of rainfall using nine natural phenomenon properties.
- Methodology: Utilization of Support Vector Machine Classifier **(**SVC) and Linear Regression (LR) models after hyperparameter tuning.
- Outcome: SVC exhibits superior performance compared to LR, both with and without hyper-tuning. Initially, the accuracy of logistic regression (LR) is 88%. After hyper-tuning, the accuracy improves to 89%. Conversely, in the case of Support Vector Machine (SVM) Classifier, the accuracy starts at 82%. After hyper-tuning, the accuracy significantly improves to 91%.

Table 1. summarizes methods used for rainfall prediction in Australia.

3. Method

SGWNN (SMOTE-GWO-Neural Network) stands out as the optimal proposed model. Figure 1 illustrates various proposed models, elaborating on their distinctions and facilitating subsequent comparative analyses.

- **Comparison Among Machine Learning Algorithms:**

Based on previous studies and utilizing the same dataset. classical machine learning techniques were employed after preprocessing and feature engineering.

Random Forest, XGBoost, SVM Classifier, and Neural Network are assessed for accuracy and evaluation errors.

- **Comparison with the SMOTE Approach:**

Evaluating the models with and without the SMOTE (Synthetic Minority Oversampling Technique) approach for enhanced performance.

- **Grey Wolf Optimization (GWO) Algorithm:**

The GWO algorithm is utilized for feature selection on the dataset. A new dataset, extracted using this algorithm, is employed with and without the SMOTE approach across machine learning models.

These comparative analyses aim to elucidate the impact of different algorithms, the influence of SMOTE, and the efficacy of the GWO algorithm for feature selection in enhancing the predictive capabilities of the models.

A) Grey Wolf Optimizer

The Grey Wolf Optimizer (GWO) is a swarm intelligence optimization technique inspired by the hunting behavior of grey wolves, introduced by Mir Jalili in 2014 [17]. This population-based metaheuristics method mimics the natural leadership structure and hunting patterns observed in Grey Wolf packs, typically consisting of 5 to 12 members on average [18]. GWO operates across four hierarchical levels, each representing a distinct role within the pack:

- **Alpha α**: Leaders within the wolf pack, both male and female, are responsible for decision-making regarding hunting, movement, and rest.
- **Beta β**: A wolf, male or female, acting as the potential replacement for Alpha. Beta assists Alpha in decision-making and provides crucial feedback.
- **Delta δ**: Wolves at this level follow and support Alpha, Beta, and Omega wolves. They fulfill roles such as sentinels, scouts, elders, caregivers, and hunters.
- **Omega ω**: The least dominant wolves, fulfilling roles of submission and often serving as scape goats. Omega wolves are crucial for following instructions from other pack members.

The GWO algorithm replicates these hierarchical roles to optimize search and decision-making processes, drawing parallels to the collaborative behaviors observed in Grey Wolf packs. In the GWO's mathematical model, the best solution is known as α. The second and third best answers were β and δ, respectively. The other additional candidate solutions ω are also expected. Three solutions are used by the GWO algorithm [17, 19]. The

mathematical model of the encircling behavior is presented in the following equations:

$$\vec{X}(t+1) = \vec{X}\, p(t) + \vec{A}.\vec{D} \quad (1)$$

where the number of iterations is t, coefficient vectors are $\vec{A}$, $\vec{C}$, the position of prey is $\vec{Xp}$, and the position of grey wolves $\vec{X}$, $\vec{D}$ is defined in equation (2).

$$\vec{D} = |\vec{C}.\vec{X}p(t) - \vec{X}(t)| \quad (2)$$

$$\vec{A} = 2a, \vec{r}1 - a \quad (3)$$

$$\vec{C} = 2\vec{r}2 \quad (4)$$

where $\vec{r}1$ and $\vec{r}2$ are random vectors in [0, 1] and a is a vector set decreasing over iterations linearly from 2 to 0. In a mathematical simulation of Grey Wolf hunting behavior, Alpha (**α**) is taken to be the best candidate for the answer, while Beta (**β**) and Delta (**δ**) are taken to have more knowledge of the potential location of the prey.

In turn, this forces others, namely (**ω**), to update their positions by the best location in the choice space by saving the three best solutions thus far. The following equations can be used to model such a hunting behavior:

$$\vec{X}(t+1) = (\vec{X}1 + \vec{X}2 + \vec{X}3)/3 \quad (5)$$

$$\vec{X}1 = |\vec{X}\alpha - \vec{A}1.\vec{D}\alpha| \quad (6)$$

$$\vec{X}2 = |\vec{X}\beta - \vec{A}2.\vec{D}\beta| \quad (7)$$

$$\vec{X}3 = |\vec{X}\delta - \vec{A}3.\vec{D}\delta| \quad (8)$$

where the positions of best solution are $\vec{X}\alpha$, $\vec{X}\beta$, $\vec{X}\delta$ which are repeated at $\vec{A}1$, $\vec{A}2$, $A3$ of equation (3), $\vec{D}\alpha$, $D\beta$, $\vec{D}\delta$ are defined as follows:

$$\vec{D}\alpha = |\vec{C}1.\vec{X}\alpha - \vec{X}| \quad (9)$$

$$\vec{D}\beta = |\vec{C}2.\vec{X}\beta - \vec{X}| \quad (10)$$

$$\vec{D}\delta = |\vec{C}3.\vec{X}\delta - \vec{X}| \quad (11)$$

where $C1$, $C2$, $C3$ are defined in equation (4) and Parameter a is updated linearly from 2 to 0 each time which is updated to control the trade-off between exploration and exploitation, defined as follows:

$$a = 2 - t\,(2/MaxIter) \quad (12)$$

where t is the number of iterations and MaxIter is the total number of iterations allowed for optimization. Finally, the GWO algorithm is expressed as follows:

Algorithm 1. Grey Wolf Optimization

- Initialize the grey wolf population Xi (i = 1, 2, ..., n)
- Initialize a, A, and C
- Calculate the fitness of each search agent
- Xα=the best search agent
- Xβ=the second-best search agent
- Xδ=the third best search agent
- **while** (t < Max number of iterations)
- **for** each search agent
- Update the position of the current search agent by equation 5
- **end for**
- Update a, A, and C
- Calculate the fitness of all search agents
- Update Xα, Xβ, and Xδ
- t=t+1
- **end while**
- return Xα

B) Machine Learning Models

In this study, a suite of machine learning algorithms has been employed for the prediction of rainfall, representing the latest advancements in this field. These algorithms have been instrumental in exploring and forecasting rainfall patterns. The selected models encompass a range of sophisticated techniques designed to address the complexities inherent in rainfall prediction.

I. Random Forest Algorithm

Random Forest (RF) is an ensemble learning method that operates by training numerous decision trees and consolidating the information from these individual trees to produce a final estimation value, typically the mean in regression tasks [20]. In comparison to single-decision trees, the Random Forest method offers a more precise calculation of error rates. Specifically, as the number of trees increases, mathematical demonstrations show that the error rate consistently converges [21]. The Random Forest algorithm offers numerous benefits, such as mitigating overfitting risks, adeptness in managing noisy or incomplete data, and handling both categorical and numerical variables seamlessly,

even with extensive datasets. Despite these advantages, a notable drawback arises in the complexity associated with interpreting and visualizing the model's outcomes and decision-making processes [22].

II. XGBoost Algorithm

The eXtreme Gradient Boost technique (XGBoost) is an ensemble learning method, specifically a rapid implementation of gradient-boosted decision trees. It has gained widespread usage, especially in recent competitions, surpassing conventional approaches and emerging as a favored algorithm in contemporary practices. Notably, XGBoost demonstrates superior accuracy compared to other algorithms, such as Random Forest. Its software implementation offers versatility, adaptability, and portability, making it particularly effective in diverse applications of gradient boosting [23].

III. Neural Networks Algorithm

Artificial Neural Networks (ANNs) represent a computational approach utilizing interconnected connections to create numerous processing units. These networks comprise cells, nodes, units, or neurons linking the input set to the output set [24]. Each node in the network is assigned a weight that signifies its relevance within the relationship. Subsequently, inputs are aggregated, and the outcome is processed through an activation function, which defines the information processing and transfer mechanisms across the network. Various activation functions exist, influencing learning types, function boundaries, variations, and network designs [25]. The architecture of an ANN delineates the model's neuron count, the layer quantity, and their interconnections. These networks consist of input, output, and hidden layers, distinguishing between single-layer and multi-layer networks based on the layer count and feedforward or recurrent networks based on the information flow direction. The learning algorithm, guided by learning paradigms, rules, and algorithm types, regulates the weights within the network [11, 26]. Despite the intricacies of the learning process, ANNs retain learned weights, contributing to their adaptability and continued optimization.

IV. Support Vector Machine Classifier

Support Vector Machines (SVMs) represent supervised machine-learning techniques applicable to classification and regression tasks. In SVMs, each feature is a coordinate value, and each data point is depicted as a point in an n-dimensional space with “n” representing the number of features. The primary goal is to identify a hyperplane that effectively separates two classes for classification purposes. SVMs establish a feature space, a finite-dimensional vector space where each dimension signifies a “feature” of an item, to model this scenario. The SVM's objective is to train a model capable of accurately categorizing new objects into predefined categories. Objects are categorized based on their position “above” or “below” the separation plane. Notably, SVMs are non-probabilistic as they lack a stochastic component, and the position of new objects in the feature space is solely determined by their features. The biased and unbiased hyperplanes involve a fraction of the training data [27].

C) Dataset

The study draws upon a dataset acquired from the Kaggle.com platform, accessible at

https://www.kaggle.com/jsphyg/weather-datasetrattle-package#weatherAUS.csv, accessed on March 10, 2022 [28]. This dataset encompasses 23 columns detailing natural phenomenon properties across 49 cities in Australia daily from 2008 to 2016. Upon data analysis, there was a significant imbalance within the dataset, with non-rain status accounting for 77.582% and 79.539% before and after data cleaning, respectively, in comparison to rain status.

Tabel 2. Machine learning models with imbalanced data

Models	Performance evaluation								
	Accuracy	MAE	MSE	RMSE	Sensitivity	Specificity	Recall	Precision	R2 Square
NN	0.865	0.13	0.13	0.366	0.506	0.957	0.506	0.755	0.17
XGBoost	0.865	0.13	0.13	0.367	0.53	0.95	0.53	0.73	0.17
Random Forest	0.861	0.138	0.138	0.37	0.458	0.96	0.458	0.77	0.147
SVC	0.865	0.13	0.13	0.367	0.48	0.96	0.48	0.77	0.17

Tabel 3. Machine learning models with balanced data

Models	Performance evaluation								
	Accuracy	**MAE**	**MSE**	**RMSE**	**Sensitivity**	**Specificity**	**Recall**	**Precision**	**R2 Square**
NN	0.907	0.09	0.09	0.3	0.87	0.94	0.874	0.938	0.63
XGBoost	0.906	0.09	0.09	0.3	0.887	0.92	0.89	0.92	0.62
Random Forest	0.91	0.08	0.08	0.29	0.897	0.92	0.90	0.92	0.64
SVC	0.905	0.09	0.09	0.3	0.867	0.94	0.87	0.94	0.6

Implementing the SMOTE (Synthetic Minority Oversampling Technique) approach is incredibly vital for these datasets. It plays a pivotal role in rectifying imbalances and ensuring a more representative and reliable dataset, especially in scenarios where class imbalance poses significant challenges for accurate modeling or prediction.

4. Result and discussion

In this section, we will discuss the high accuracy achieved by the proposed model of 99.89%. Furthermore, the performance of various machine learning models with overbalanced and imbalanced datasets, employing the SMOTE approach, will be analyzed. This study delves into the impact of Grey Wolf Optimization on these models and compares their performance with the proposed model, providing comprehensive insights into the efficacy of different methodologies in predicting rainfall.

- **Comparison Between Used Models without The SMOTE Approach**

In this section, the performance evaluation of various models in predicting rainfall using imbalanced data is presented in Table 2. Among these models, Neural Network (NN) emerges as the most effective in terms of predictive accuracy. This comparison highlights the superior performance of NN over other models when considering imbalanced datasets for rainfall prediction.

- **Comparison Between Used Models with The SMOTE Approach**

In this section, the performance assessment of models in predicting rainfall using balanced data is outlined in Table 3. Among the evaluated models, Random Forest emerges as the top-performing model. This comparison underscores Random Forest's superior performance when applied to balanced datasets for rainfall prediction, following the SMOTE approach. Comparing the results outlined in Tabel. 2,3 with those mentioned in Tabel. 1, we attained higher accuracy rates with the Neural Network (NN) model. Specifically, we achieved an accuracy of 86.5% with imbalanced data and 90.7% with balanced data, whereas previous studies reported 84% accuracy with the same model [11]. With the Random Forest (RF) model, we achieved accuracy rates of 86.1% with imbalanced data and 91% with balanced data, while [11] achieved 90.99%. Similarly, with the XGBoost model, we obtained accuracy rates of 86.5% with imbalanced data and 90.6% with balanced data, compared to 90.46% reported in [11]. These findings indicate that the models performed exceptionally well with balanced datasets. However, despite these promising results, we were not entirely satisfied and thus sought to optimize them further by incorporating the Grey Wolf Optimization (GWO) algorithm in the subsequent section.

- **Comparison Between Used Models with GWO Algorithm.**

This study explores the application of the GWO algorithm for effective feature selection [GWO- (NN, RF, XGBoost, SVC)] in predicting rainfall status. The GWO algorithm is initially utilized to eliminate redundant and irrelevant features by creating initial positions in the discrete search space and updating the population's positions iteratively. Parameters such as the number of wolves (5), iterations (100), dimensions (113), search domain ([0,1]), $\alpha \in [0,1]$, and $\beta = 1 - \alpha$ are used for feature selection, resulting in the extraction of 76 pertinent features. Upon feature selection, it is observed that the dataset maintains a notable imbalance, with a 79.539% representation of “no rain” status, suggesting the potential application of the SMOTE approach. Subsequently, machine learning models are operated using these optimal features to predict rainfall in Australia based on the 113 meteorological properties.

The ensuing table in Table 4 discusses the accuracy of these models in predicting rainfall based on the optimal feature set obtained through the GWO algorithm, providing insights into their respective performances.

Tabel 4. Performance of machine learning-based GWO

Models	Accuracy with GWO	
	Accuracy with Imbalanced Data	**Accuracy with Balanced Data**
GWO-NN	**86.3%**	**99.89%**
GWO-RF	86.14%	99.968%
GWO-XGBoost	86.2%	99.975%
GWO-SVC	86.26%	99.979%

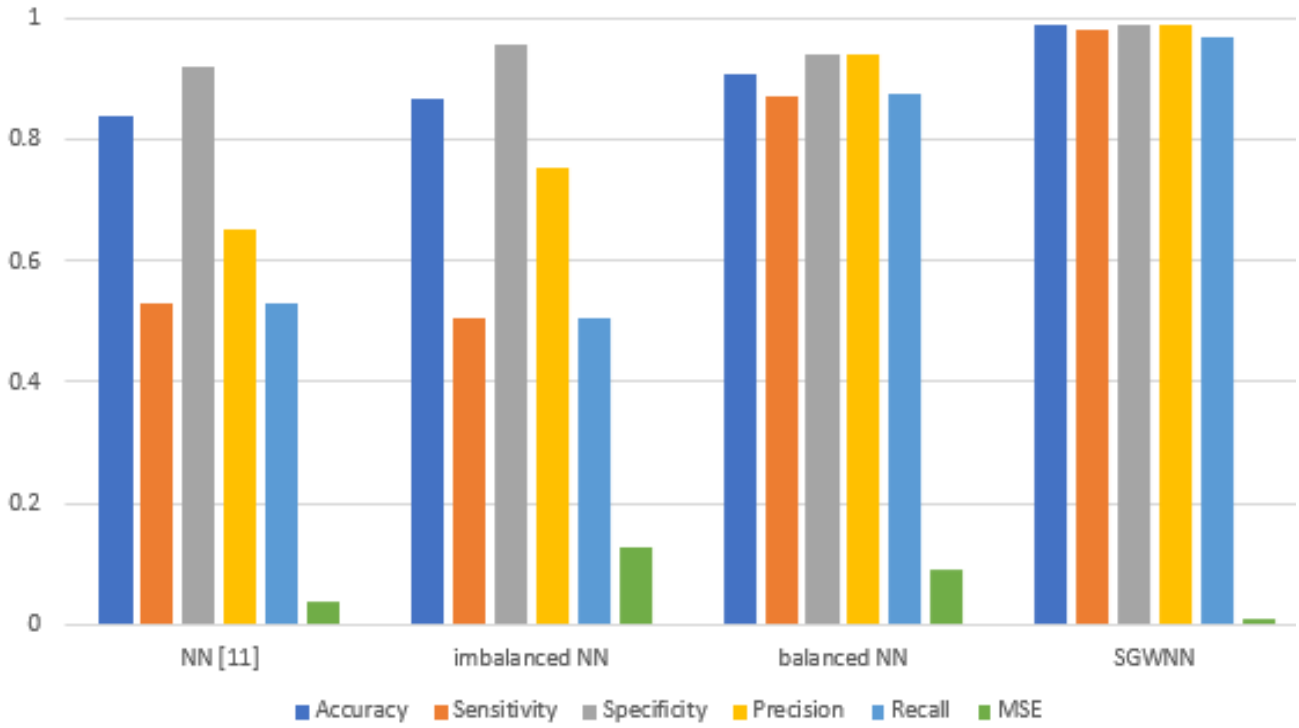

Figure. 2 Comparative Analysis of Techniques Using Neural Network

The previous results demonstrate that the Neural Network (NN) applied to the optimized feature selection derived from the Grey Wolf Optimization (GWO) algorithm showcases improved accuracy when complemented with the SMOTE (S) approach. Consequently, this experimental analysis confirms that the proposed SGWNN model stands out as the most effective model, achieving a notably high accuracy of 99.89% in predicting rainfall within this study. We observed a strong correlation between the Neural Network (NN) and Grey Wolf Optimization (GWO). The method proposed in this study consistently demonstrated significantly higher accuracy compared to other machine learning methods as showing in Figure 2.

5. Conclusion

In this study, Recent observations suggest that combining GWO and Neural Network for rainfall prediction in Australia yields stronger correlations, particularly when applied to balanced data that (referred to as the SGWNN model). Previous models struggled due to data complexity and the imbalance between rain and no-rain statuses. SMOTE balanced the data, highlighting Random Forest as the most accurate model at 91% accuracy. Our goal was to enhance models' performance by employing GWO for feature selection, showing significant improvements. Combining SMOTE with GWO in Neural Networks offers significant advantages: it

addresses imbalanced data by creating a balanced dataset, aiding in a fair representation of different classes. This fusion enhances model performance by refining parameters, reducing bias, and improving accuracy against imbalances. It improves the network's ability to handle new data by reducing bias and ensuring a balanced learning process. Overall, this approach creates more reliable predictions, which is especially crucial for underrepresented minority classes. The SGWNN model notably achieved 99.89% accuracy. Our study demonstrates that SGWNN model are more resilient than previous models. Moving forward, further investigations spanning data from 2019 to the present across Australia and other countries can expand our understanding of the SGWNN model's robustness and applicability in diverse meteorological contexts.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

“Conceptualization, Kamal ElDahshan and H. Mancy; methodology, Amira ElKhateeb; software, Amira ElKhateeb; validation, H. Mancy, Amira Elkhateeb, and Kamal ElDahshan; formal analysis, Amira Elkhateeb; investigation, H. Mancy; resources, Amira ElKhateeb; data curation, Amira Elkhateeb; writing—original draft preparation, H. Mancy; writing—review and editing, Amira Elkhateeb; visualization, H. Mancy; supervision, Hoda A. Ali; project administration, Kamal ElDahshan and Hoda A. Ali; funding acquisition, H. Mancy.

Acknowledgments

This study is supported via funding from the Prince Sattambin Abdulaziz University (project number PSAU/2024/R/1445).

References

[1] A. Datta, S. Si, and S. Biswas, "Complete Statistical Analysis to Weather Forecasting", *International Journal of Computational Intelligence in Pattern Recognition, Advances in Intelligent Systems and Computing*, Springer, Singapore, Vol. 999, pp. 751–763, 2020, doi: 10.1007/978-981-13-9042-565.

[2] A. Elkhateeb, H. Mancy, M. Zaki, and K. Eldahshan, "Dependency of the learning technique on the problem nature", *International Journal of Theoretical and Applied Research*, Vol. 2, No. 1, pp. 77-84, 2023, doi: 10.21608/ijtar.2023.142717.1010.

[3] A.Y. Barrera-Animas, L. Oyedele, M. Bilal, T. Akinosho, "Rainfall Prediction: A Comparative Analysis of Modern Machine Learning Algorithms for Time-series Forecasting", *International Journal of Machine Learning with Applications*, Vol. 7, pp. 100-204, 2022, doi: 10.1016/j.mlwa.2021.100204.

[4] P. Abbaszadeh, K. Gavahi, and H. Moradkhani, "Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting", *International Journal of Advances in Water Resources*, Vol. 145, pp. 103721, 2020, doi: 10.1016/j.advwatres.2020.103721.

[5] R. Mumtaz, S.M.H. Zaidi, M.Z. Shakir, U. Shafi, M.M. Malik, A. Haque, S. Mumtaz, and S.A.R. Zaidi, "Internet of Things (IoT) Based Indoor Air Quality Sensing and Predictive Analytic—A COVID-19 Perspective", *International Journal of Electronics*, Vol. 10, pp. 184, 2021, doi: 10.3390/electronics10020184.

[6] Y. Ohashi, T. Ihara, K. Oka, Y. Takane, Y. Kikegawa , "Machine learning analysis and risk prediction of weather-sensitive mortality related to cardiovascular disease during summer in Tokyo, Japan", *International Journal of Sci Rep*, Vol. 13, p. 17020, 2023, doi: 10.1038/s41598-023-44181-9.

[7] M. Mageshwari, R. Naresh, "Improved Sunflower Optimization Algorithm Based Encryption with Public Auditing Scheme in Secure Cloud Computing", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 6, 2023, doi: 10.22266/ijies2023.1231.02.

[8] B. Singh, Y.D.S. Arya, K.C. Tripathi, "Deep Learning Analysis of Impact of Arctic Sea Extent Over Indian Ocean Sea Surface Temperature", *International Journal of Intelligent Engineering and Systems*, Vol. 16, No. 6, 2023, doi: 10.22266/ijies2023.1231.01.

[9] E.A. Hussein, M. Ghaziasgar, C. Thron, M. Vaccari, and Y. Jafta, "Rainfall Prediction Using Machine Learning Models: Literature Survey", In: *Proc. of International Conf. Artificial Intelligence for Data Science in Theory and Practice*, Vol. 1006, 2022, doi: 10.1007/978-3-030-92245-0_4.

[10] D.H. Nguyen, J.-B. Kim, and D.-H. Bae, "Improving Radar-Based Rainfall Forecasts by Long Short-Term Memory Network in Urban

Basins", *International Journal of Water*, Vol. 13, p. 776, 2021, doi: 10.3390/w13060776.

[11] A. Sarasa-Cabezuelo, "Prediction of Rainfall in Australia Using Machine Learning", *International Journal of Information*, Vol. 13, p. 163, 2022, doi: 10.3390/info13040163.

[12] R.O. Imhoff, C.C. Brauer, K.J. van Heeringen, R. Uijlenhoet, and A.H. Weerts, "Large-sample evaluation of radar rainfall nowcasting for flood early warning", *Water Resources Research*, Vol. 58, 2022, doi: 10.1029/2021WR031591.

[13] Z. He., "Rain Prediction in Australia with Active Learning Algorithm.", In: *Proc. of International Conf. on Computers and Automation (CompAuto)*, Paris, France, pp. 14-18, 2021, doi: 10.1109/CompAuto54408.2021.00010.

[14] T. Shah, "POM: Predicting Rain Using ML-DL Tech", *Kaggle*, [Online]. Available: https://www.kaggle.com/code/twaritshah/pom-predicting-rain-using-ml-dl-tech/notebook

[15] A.M. Kulkarni, S.P. Tidake, M.V. Shelke, A.L. Devkar, and V.A. Deshmukh, "WEATHER AUS DATASET: PREDICTION OF RAINFALL USING MACHINE LEARNING TECHNIQUES", *International Journal of Eur. Chem. Bull*, Vol. 12, No. S3, pp. 1531-1538, 2023, doi: 10.31838/ecb/2023.12.s3.1702023.19/04/2023.

[16] G. Tuysuzoglu, K. Ulas Birant, D. Birant, "Rainfall Prediction Using an Ensemble Machine Learning Model Based on K-Stars", *International Journal of Sustainability*, Vol. 15, No. 7, p. 5889, 2023, doi: 10.3390/su15075889.

[17] F. Aderyani, S. Jamshid Mousavi, F. Jafari, "Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN", *International Journal of Hydrology*, Vol. 614, p. 128463, 2022, doi: 10.1016/j.jhydrol.2022.128463.

[18] R. Mohd, M. Ahmed Butt, M. Zaman, "Grey Wolf-Based Linear Regression Model for Rainfall Prediction", *International Journal of Information Technologies and Systems Approach*, Vol. 15, No. 1, pp. 1-18, 2021, doi: 10.4018/ijitsa.290004.

[19] M. Karimi-Mamaghan, M. Mohammadi, P. Meyer, A. Mohammad Karimi-Mamaghan, E. Talbi, "Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art", *European Journal of Operational Research*, Vol. 296, No. 2, pp. 393-422, 2022, doi: 10.1016/j.ejor.2021.04.032.

[20] M.S. Balamurugan and R. Manojkumar, "Study of short-term rain forecasting using machine learning based approach", *International Journal of Wireless Networks*, Vol. 27, pp. 5429-5434, 2021, doi: 10.1007/s11276-019-02168-3.

[21] R. Praveena, T. R Ganesh Babu, M. Birunda, G. Sudha, P. Sukumar, J. Gnanasoundharam "Prediction of Rainfall Analysis Using Logistic Regression and Support Vector Machine", *International Journal of Physics: Conference Series*, Vol. 2466, No. 1, p. 012032, 2023, doi: 10.1088/1742-6596/2466/1/012032.

[22] M. Irshad and V. Kumar, "SMOTE and Extra Trees Regressor based random forest technique for predicting Australian rainfall", *International Journal of Inf. Tecnol.*, Vol. 15, pp. 1679–1687, 2023, doi: 10.1007/s41870-023-01185-y.

[23] M.T. Anwar, E. Winarno, W. Hadikurniawati, M. Novita, "Rainfall Prediction Using Extreme Gradient Boosting", *International Journal of Physics: Conference Series*, Vol. 1869, No. 1, p. 012078, 2021, doi: 10.1088/1742-6596/1869/1/012078.

[24] S.K. Dewangan, C. Nagarjuna, R. Jain, R.L. Kumawat, V. Kumar, A. Sharma, B. Ahn, "Review on applications of artificial neural networks to develop high entropy alloys: A state-of-the-art technique", *International Journal of Materials Today Communications*, Vol. 37, p. 107298, 2023, doi: 10.1016/j.mtcomm.2023.107298.

[25] O. Altay and E. Varol Altay, "A novel hybrid multilayer perceptron neural network with improved grey wolf optimizer", *International Journal of Neural Comput & Applic*, Vol. 35, pp. 529–556, 2023, doi: 10.1007/s00521-022-07775-4.

[26] M.A. Millán, A. Picardo, R. Galindo, "Application of artificial neural networks for predicting the bearing capacity of the tip of a pile embedded in a rock mass", *International Journal of Engineering Applications of Artificial Intelligence*, Vol. 124, pp. 106568, 2023, doi: 10.1016/j.engappai.2023.106568.

[27] C. Atik, R.A. Kut, R. Yilmaz, D. Birant, "Support Vector Machine Chains with a Novel Tournament Voting", *International Journal of Electronics*, Vol. 12, p. 2485, 2023, doi: 10.3390/electronics12112485.

[28] Kaggle. Weather Dataset (Rattle Package). [Online], 2021, Available: https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package