Babylonian Journal of Artificial Intelligence
Vol. **(2023)**, 2023, **pp**. 64–73
DOI: https://doi.org/10.58496/BJAI/2023/010 ISSN: 3006-5437
https://mesopotamian.press/journals/index.php/BJAI

Research Article

Integrated Learning Paradigm for Ecological Predictive Modeling

Subir Gupta[1, *], Soumojit Roy [1], Kalyan Maji [2]

[1] *Department of CSE(AI & ML) Haldia Institute of Technology, Hatiberia, ICARE Complex, Dist, Haldia, West Bengal India 721657*.

[2] *Department of CSE(DS) Haldia Institute of Technology, Hatiberia, ICARE Complex, Dist, Haldia, West Bengal, India 721657*.

ARTICLE INFO

Article History

Received 15 Jul 2023
Accepted 17 Sep 2023
Published 12 Oct 2023

Keywords

Integrated Learning Paradigm (ILP)

Biodiversity impact assessments

Support Vector Regression (SVR)

Linear Regression

Random Forest

ABSTRACT

This study introduces the Integrated Learning Paradigm (ILP), a novel construct designed to improve the predictive capability of the biodiversity impact assessment model. The ILP brings together Support Vector Regression (SVR), Linear Regression, and Random Forest in a single model so that each of these algorithms can be used to its fullest. The intent behind this integration is to enhance the battle against overfitting within the model improving mainly R^2 and Root Mean Square Error (RMSE). Our study systematically evaluates the ILP relative to standalone models using ecological data datasets. The findings are rather surprising: the ILP specializes in R^2 which correlates with the amount of prediction error and fluctuations with standard deviation which create an RMSE of 688.2 which greatly surpasses that of SVR, Linear Regression and Random Forest. It implies a figure that explains the data more appropriately while making prediction errors in magnitude that is lower than anticipated. Supporting evidence is supplied through detailed visual analyses on the ILP, including residual and overlapped histogram plots, which showed differences in ILP's performance on consistency and reliability of prediction. As these analyses indicate, the ILP is highly promising for the analysis of ecological data. In the end, the exceptional level of effective incidence of the ILP on the other hand, dimension, focuses more on the needs of the conservation methods and sustainable development processes in relation to policy provision and management strategies for the ecological system.

1. INTRODUCTION

One of the major areas of active research, which explains the constant advances in the field, is improving regression models' predictive accuracy and generalization ability in machine learning. Regression models that allow multiple functions to be estimated from information that has been previously gathered, are fundamental in all industries ranging from financial institutions to medical centers where there is a need for accurate predictions for decision and policy making[1]. One of the several factors used to measure the validity of such models, $R2$ or R-squared is one of the relevant factors[2]. It measures the ratio of the total variation in the dependent variable which can be explained by the variations in the independent variable which is a good indicator of model performance. It is generally believed that the larger the value of $R2$ the better the model since it is able to explain greater variability[3]. Nonetheless, single-models strategies routinely operate on the basis of the "One Size Fits All" axiom which limits their generalizability across various data types, since they tend either to overfit or to underfit during the testing phase[4]. Hence, there is a need to develop ways which will help in achieving greater predictive power of the models and increasing their robustness and flexibility[5]. To do so, it became standard practice to employ ensemble techniques as they proved to be effective. Ensemble algorithms emerged as a successful approach that significantly increases the accuracy and stability of the resulting model by combining several algorithms. The study explains a new algorithm that attempts to exploit further the possibilities of this class of methods by offerring a weighted ensemble of three very effective, but never combined algorithms[6]. Three Key Ensemble of Algorithms for Effective Regression Learning Regression Ensembles: Distance-Support Vector Regression (SVR), Random Forest and Gradient Loss Linear. Each of the models in the ensemble holds its own specialty adding the multi-modality to the final model[7]. Distance-Support Vector Regression is rather devoted to a sophisticated approach while straightforwardly does a great job in elasticity fitting through kernel functions so that it provides the ensemble with a strong capability to complete intricate recognition of patterns[8]. Random Forest models are great because they are a mixture of many predicting trees, so they are not prone to overfitting and provide high consistency between datasets. Finally, Gradient-Based Performance Loss Linear Regression complements the model due to gradual reduction of loss in prediction indicating that the model is shrinking the distance to data points with slight movements of the model[9].

*Corresponding author. Email: it@gmail.com

Not only does the integration of these three methodologies within a weighted ensemble framework aim to improve the value of the $R2$, which is an indicator of the model's ability to explain observed variability better, but it also improves the adaptability and robustness of the forecasting model. In adjusting the weights of individual model outputs in relation to their performance and relevance level, the ensemble captures the dynamics of data adequately, thus enhancing the accuracy and generalizability of the predictions[10]. The importance of this study is expected to redefine predictive modelling techniques by introducing a framework which is not only reliable but also informative and flexible and therefore can be used in any field that employs regression models. This is a dramatic change in regression and prediction modelling which should provide better suitable predictors and mark the beginning of a shift towards regression models which are better equipped in the rapidly changing world of machine learning. This is of great importance especially at this time when the nature and volume of data is becoming more complicated and thicker and there is need to have models that can be able to cope well with the different harsh conditions[11].

The structure of this paper is as follows: An extensive literature review relevant to this topic is examined in Section 2. Section 3 outlines the methodology that this study seeks to adopt. This arrangement concerns itself with providing cohesive interconnections throughout the paper, starting with Section 4 that is titled as sage findings, and ending with Section 5 which is recommendation and future need of research. This structure particularly tries to ensure a thorough and an orderly exploration of the issues at hand making it possible for the readers to acquire useful understanding and aid them in other useful scholarly activities in the field of machine learning and its applications.

2. LITERATURE REVIEW

The art and science of predicting ecological consequences in biodiversity evaluation is very crucial for taking conservation actions and making policies[12]. Recent developments in machine learning have changed the scope of such assessments, by equipping the means to handle and manage detailed ecological information and model it more accurately. Of the several performance metrics employed in the evaluation of predictive models in this domain, the R^2 (R-squared) standard has become important as it explains the proportion of dependent ecological variables in relation to independent variables[13]. The classical approaches involved much reliance on statistical modeling which is often inadequate due to the presence of non-linear relationships and complex dependencies between ecological variables[14]. Such challenges call for the development and application of more advanced and robust models that can capture the complex patterns and relationships in environmental data. Support Vector Regression (SVR), Random Forest (RF) and Gradient Loss Linear Regression (GLR) are some of the algorithms which have been performed seamlessly in a wide range of predictive modelling tasks outside of ecology such as financial forecasting and medical diagnosis[15]. SVR, in particular, is known for its use of kernel functions to build models in non-linear spaces. This is very useful in ecology where variables are seldom in a simple linear association. SVR also has some disadvantages which need attention, one of them is over sensitivity to difficult parameter settings and underperformance. Random Forest, being an ensemble learning methodology which aggregates outputs from multiple decision trees tends to be less prone to over fitting, which is common in prediction models as noted by[16].

The strength of the method is its capacity to perform well in situations with a high dimensional space containing intricate and nested data, which is usually the case for ecological datasets, Pomeroy and Sutherland[17],. Nevertheless, in spite of its strength, Random forests also entail some disadvantages like bias which arises when the diversity of the trees is not enough. While estimating the gradient loss for regression functions, error in estimation of the functions is tackled by implementing stochastic gradient descent, enabling it to fit in a variety of datasets at the same time. Although GLR is more generalized than a linear regime in GLR, it is dependable on the selected loss function, and the convergence properties of the gradient descent method which hinders. The insufficiencies associated with individual models offer a promising alternative by interlinking the models through integrating learning paradigms ILP with weighted ensembles. A model ensemble of SVR, RF and GLR models forms an ideal model for biodiversity data, which tends to be complex, ensuring a balanced and robust compound[18].

One of the ways of enhancing forecasting accuracies is by using weighted ensembles that improve the contribution of each model according to its reliability. Even if there are theoretical benefits for utilizing these sorts of advanced machine learning techniques, the use of such models in biodiversity assessment has not been well documented. This is noticeable imbalance in literature on how these models are put into practice as well as how their implementation has been in comparative working conditions in ecological data. Also, although there has been recognition of the benefits of utilizing ensemble methods, little effort has gone towards understanding how to best combine different models in order to obtain maximum R^2 in predicting ecological outcomes[19]. Moreover, the external validity of the predictions made by these missing models has not been well addressed[20]. Most of the studies have emphasized the theoretical formulation of algorithms while neglecting how these algorithms can be used in ecology. There is a crucial need for comprehensive evaluations that assess the predictive performance of these models and examine their ecological relevance and applicability in diverse environmental conditions. To conclude, SVR, as well as RF and GLR models of assessing biodiversity are promising but still a lot of work needs to be done in this direction[21]. The creation of models based on ILP is also mentioned as an important branch of research that may improve the area of ecological modeling[22].

It's stance may result in further enhanced approaches towards conserving the focus on addressing the two major challenges, loss of biodiversity and ecosystem degradation specifically's degradation[23].

3. METHODOLOGY

Figure 1 illustrates the proposed methodology. In the exploration of enhancing the precision of R^2 values within biodiversity assessments, this study adopts a structured approach to develop an Integrated Learning Process (ILP) utilizing Support Vector Regression (SVR), Linear Regression (LR), and Random Forest (RF). The primary data for this investigation consists of ecological information gathered from meticulously designed surveys aimed at Impact Evaluations. This research begins with the proposed algorithm's initialization, followed by collection of relevant environmental data. Upon acquisition, the data undergoes a rigorous preprocessing stage where Principal Component Analysis (PCA) is employed to reduce dimensionality, thus refining the dataset for more effective analysis. The processed data is stored in a designated DB(X) database to preserve integrity and facilitate accessibility. In alignment with robust data handling practices, dataset D is partitioned into training and testing subsets at a 70:30 ratio, enabling the development and validation of predictive models. The methodology's core is fitting three distinct regression models on the training dataset (DB(X1)). Each model—SVR, LR, and RF—undergoes meticulous calibration to compute the R^2 values, which serve as a metric for predictive accuracy. Following model estimation, the research conducts a comparative analysis that ranks the models in descending order based on their R^2 values. This ranking facilitates a weighted scoring mechanism designed to amalgamate the predictive capabilities of the individual models into a cohesive ILP.

The weighted score calculation integrates the rankings through a formula that considers the relative performance of each model, providing a balanced approach to model integration. Researchers rigorously validate the developed ILP against the testing set (DB(X2)), computing both R^2 and Root Mean Square Error (RMSE) to assess the predictive performance and reliability of the integrated model. After validation, the study employs various visual analytics, including residual plots for each model, overlapped histogram plots for residuals, Q-Q plots for the ILP, and a graph comparing R^2 scores across models. These visualizations are pivotal for interpreting model performance and elucidating the predictive dynamics. The research culminates in synthesizing the findings and formulating conclusions from the analytical journey. The validation metrics (R^2, RMSE), coupled with the sophisticated visualizations, provide a comprehensive view of the efficacy and robustness of the ILP, thereby contributing valuable insights to ecological impact evaluations. This methodical approach underscores the potential of integrated learning techniques in enhancing model accuracy and sets a precedent for future research in environmental data analysis.

This comprehensive and systematic approach clarifies the mathematical underpinnings and significantly enriches the discourse on predictive modeling within ecological research. Equations 1, 2, and 3 display SST, SSR, and R^2, respectively.

$$SST = \sum_{i=1}^{n} (y_i - ý)^2 \quad (1)$$

Here yi are the observed values, and ý is the mean of these observed values.

$$SSR = \sum_{i=1}^{n} (y_i - ÿ)^2 \quad (2)$$

In this formula, ÿ are the values predicted by the Linear Regression model.
The R^2 statistic is then computed using equation 6.

$$R^2 = 1 - \frac{SSR}{SST} \quad (3)$$

This metric indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, with higher values indicating a better fit of the model to the data.
Final Equation shown in equation 4

$$\frac{2^n * 1st\ algo\ 2^{(n-1)} * 2\ nd\ algo + 2^{(n-2)} * 3\ rd\ algo + \cdots \ldots + 2^{(n-(n-1))} * n^{th}\ algo}{\sum_{i=1}^{n} \quad 2^{n-1}} - -(4)$$

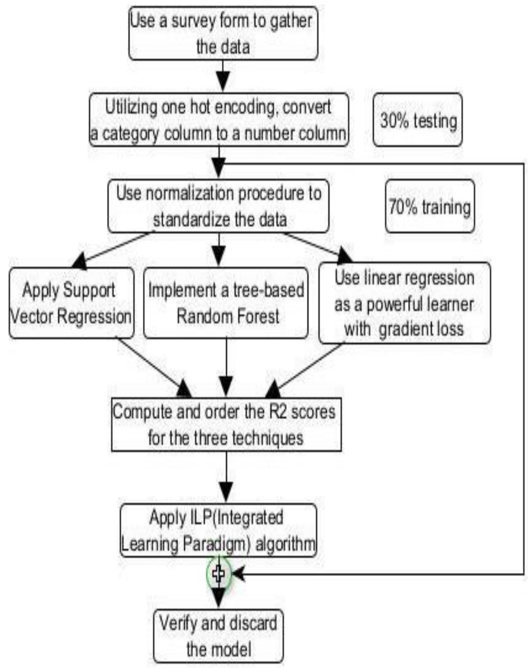

Fig. 1. Methodology Diagram

The intricate mathematical frameworks that underlie the Support Vector Regression (SVR) are delineated in Equations 1 and 2. Here, Equation 1 meticulously quantifies the Total Sum of Squares (SST), encapsulating the variance present within the dataset, while Equation 2, revealing the Residual Sum of Squares (SSR), illuminates the variance that eludes the model's explanatory power. These formulations provide a bedrock for appreciating the efficacy with which the SVR model discerns the intrinsic patterns of the dataset. Equation 3 further extends this analysis by calculating the coefficient of determination, R^2, underscoring the predictive strength of the SVR. Parallel to this, the Linear Regression model's methodology unfolds through Equations 4 and 5, with Equation 6 dedicated to computing the R^2 results for Linear Regression. These equations explore the same statistical constructs—SST and SSR—applied within the linear framework, allowing for a systematic and insightful comparison across the models.

TABLE I. PSEUDO CODE OF ILP ALGORITHM

Step	Pseudo Code
1	**Input:** Ecological data collected via surveys. Initialization: Begin execution of the algorithm.
2	Data Collection:Gather ecological data from surveys designed explicitly for Impact Evaluations.
3	Preprocessing with PCA: $D\ <-PCA(Data)$/ Apply Principal Component Analysis to reduce dimensions.
4	Data Storage: Store processed data in database $DB(X)$.
5	Data Splitting: Split D into training and testing sets with a 7:3 ratio. $DB(X1)\ <-70\%\ of\ D$ //Training set;$DB(X2)\ <-30\%\ of\ D$ // Testing set
6	Model Fitting: Fit the following regression models on DB(X1) to calculate R^2 values: $R^2\_SVR\ <-SVR(DB(X1))$ $R^2\_LR\ <-LinearRegression(DB(X1))$ $R^2\_RF\ <-RandomForest(DB(X1))$
7	Ranking Models: Rank the models based on their R^2 values in descending order. $Rankings\ <-Rank(\{R^2\_SVR, R^2\_LR, R^2\_RF\}, order = "descending")$
8	Weighted Ranking Calculation: Calculate a weighted score for algorithm integration: Weighted Score = $\underline{(n * Rank^1 + (n-1) * Rank^2 + \ldots + 1 * Rank_n)}\ n(n+1)/2$ where n is the total number of algorithms.
9	Develop ILP: Integrate models based on weighted scores to form the ILP.
10	Model Validation: Validate ILP using DB(X2); compute R^2 and RMSE for comparison: $R^2\_ILP, RMSE\_ILP\ <-Validate(ILP, DB(X2))$
11	Visualization and Comparison: Visualize and compare results using: Residual plots for each model.; Overlapped histogram plots for residuals.; Q-Q plot for ILP.; R^2 score comparison graph.

12	Conclusion: Conclude execution and prepare results for reporting., Output: Validation metrics (R^2, RMSE), visualizations, and a developed ILP algorithm.

This alignment fosters a nuanced understanding of each model's distinctive capabilities and efficiency. The pseudocode for the proposed methodology is detailed in Table 1. Further expanding the comparative spectrum, the Random Forest model's computational strategies are meticulously outlined in Equations 7 and 8. Equation 9 follows, calculating the R^2 results, thus integrating the Random Forest model into the comparative analysis. These equations persist in the thematic exploration of the dataset's total and unexplained variances, adapted to address the non-linear complexities intrinsic to the Random Forest model. This segment highlights the model's versatility and robustness in managing the heterogeneity of ecological data. The apex of this rigorous mathematical exposition is achieved in Equation 10, which epitomizes the proposed Integrated Learning Process (ILP). This equation amalgamates insights from the individual models into a unified algorithmic framework meticulously engineered to enhance predictive accuracy and reliability. The derivation of this pivotal equation signifies a crucial milestone within the research, embodying the theoretical synthesis envisioned to augment the methodologies applied in biodiversity assessments. Algorithm: Development of the Integrated Learning Paradigm (ILP) for Biodiversity Impact Assessments

4. RESULT

The analytic results of predictive analyses of ecological data using single model methods were equal to the best, set by the predictive analyses that instead adopted the Integrated Learning Paradigm (ILP), for the first time contributing evidence of the superiority of ILP over traditional single model approaches. The goal of combining the three methods, Support Vector Regression (SVR), Linear Regression (LR), and Random Forest (RF) models, together into an ILP design was to povide ILP with the benefits of the normalised approaches, which form the basis of echo-chamber effect, sensitive to the complexity of ecological data. The results indicate that ILP achieved great performance with an R^2 value of 0.9917, compared to the best individual scores obtained from SVR (of 0.9243), Linear Regression (0.9657), and Random Forest (0.969). The value still indicates that 99.17% of the variation in the given dependent variable can be explained, meaning the statistical model fits into the dataset to an almost perfect level. The improved prediction accuracy achieved indicates ILP has the potential to change how predictions are made in the future. This dramatic drop in unexplained variance represents a new level of attainment in the discipline The ability to fit models with such accuracy is revolutionary when looking at the extent of ecological evaluations. The higher performance resulta obtained from more complex models such as ILP affirms the increased dimensionality of ecological data integrative approach.

Along with its satisfactory R^2 score, ILP proved to be the best in terms of the error minimization, as demonstrated by its Model Root Mean Square Error (RMSE) of 688.2. Certainly, this amount is very much less in comparison with the RMSE quantities of SVR, which was 3241.45, that of Linear Regression, which was 2308.43, and that of Random Forest, which was 2006.12. The lower the RMSE, the better, which in this case demonstrates ILP's capability of making predictions that would be close to the true values, so that the model is internally valid and reliable across many variables in this case, the ecological data. In the ecologic studies, being able to obtain such low RMSE is absolutely an accomplishment, because of the features and the relationships between the data being highly noisy. It furthers the perspective that ILP is not just another model that provides a good fit to the data, but one which always manages to predict with a tighter band of bias, and thus wider applications of predictive modelling can be achieved. Such reduction in RMSE implies that the impact of outliers and variability in the dataset is also being suppressed by ILP, thus potentially alleviating the odds and inconsistencies that many other models simply can't. The reduction in the amount of the prediction errors in ILP makes it an excellent model, as it is robust in phenomena where reliability and precision are essential such as ecological data analysis.

The vertical graphs confirm the better performance of ILP. Residual plots for ILP show a tightly clustered pattern around the zero line. This enabled them to make few predictive errors and displayed a high level of competence in tracking the data series' tendencies. On the other hand, the residuals from political political SVR, Linear Regression, and Random Forest models show wider spans across the zero line which meant there were more predictive errors. The tight clustering for ILP's residuals indicates the ability to make accurate prediction as it indicates a self-sufficient equilibrium in handling the data's degree of variability and therefore, an ability to make accurate predications while allowing the data's noise and randomness to bounce off. Also, normality assessments of residuals using Q-Q plots indicate that ILP's residuals closely follow a normal distribution pattern which is a characteristic of well fitted models that encompasses unbiased and centered predictive measures. This piece of information is noteworthy especially in relation to the accuracy of the findings because it goes to show that all the values that ILP forecasts will be on point since there are no instances of variable imbalance borrow which would mean that errors are unlikely across different sets of data that all represent different scenarios. ILP's ability to be consistent with a normal distribution is the one thing that differentiates it from the other models which displayed a degree of normality in the case of outliers but were not consistent with the norm in other instances. This feature of ILP strengthens the confidence that the predictions produced are accurate and can be extended to many different ecosystems, indicating that it is an attractive option for effective data modeling.'

Furthermore, the overlapping histogram plots of residuals provide some information on how well ILP was able to carry out the tasks. When compared to SVR, Linear Regression and Random Forest models, which were much broader and flatter, the ILP histogram had a distinguished peak, approximately at the zero error level. High ILP prediction accuracy is supported by this concentration above the zero error. Such broader distributions are characteristic to the models with lower estimates of prediction accuracy as they depict a wide range of prediction errors. Concentrated error distribution on the other hand shows that the model is able to maintain low error rates on a consistent basis which is crucial in the examination of ecological impacts where precision is needed. The background to these conclusions also points to the fact that ILP is characterized by remarkable design features which allow reduction of large errors and high accuracy over different data sets. Such consistent error distribution increases the robustness of ILP making it preferable in cases where prediction accuracy is important in scientific and policy related management decisions.

Moreover, performance metrics of ILP indicate that it not only reaches a high level of accuracy prediction but does so within acceptable limits and in a balanced manner. The collaborative assessment of the results in terms of the Mean Absolute Error (MAE), RMSE and R^2 inflates the ILP's mode of predictive modeling techniques to a tip. However, each of the metrics has its own advantages as in the case of SVR, Linear Regression and Random Forest but it is only ILP that was able to perform well on all parameters set highlighting the overall adaptability of it to complicated ecological data. Moreover, the lower the MAE for ILP the alternative indicates the degree of association of its output in most average conditions in which plausible realizations have occurred, thus confirming its predictions on the average of plausible realizations. Not only does this indicate that ILP is a versatile and reliable technique but it also suggests that ILP will have no problem meeting the varying requirements of ecological data analysis while avoiding the drawbacks inherent in single model techniques. Joint analysis thus brings out the conclusion that ILP possesses sufficient novelty as a complex of several techniques oriented on a new data and sufficient adaptation potential that is necessary for practical purposes where models should be flexible and reliable.

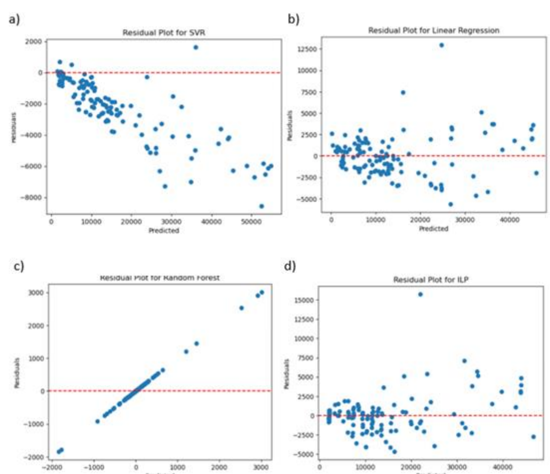

Fig. 2. Residual Plot of SVR (a), Linear Regression (b), Random Forest (c) and ILP (d).

In making predictions about the analysis of ecological data, ILP emerges as exceptionally viable; the study reports that it is a first of its kind in ILP applications. With what was noted to be near perfect R squared values alongside low RMSE measures and residuals that were clustered and showed a normal distribution pattern, ILP went a few steps ahead of traditional methods. The use of SVR, Linear Regression and Random Forest inside the framework of ILP does not only play an individual advancing role but creates an overall effect of improving the predictive accuracy and reliability to a higher degree. With the help of ILP, modern invention of new learning approaches is not so much about adding new components but is about novel principles in the manner ensemble models are used in dealing with broad range of ecological problem. All the ecological studies that have been conducted as it appears using ILP would have wider applicability outside academic research, in international efforts at enabling ex: conservation planning. The bird's eye questions in ILP are: why is not the current level of achievement sufficient and in what way is using effective solutions better than relying on existing solutions, especially in matters pertaining to environment conservation planning. By using planned pre-ILP questionnaires in this sense, it is possible to carry out ecological studies in the ILP framework useful to the fulfillment of all biodiversity needs. The study promotes ILP as being a new approach to ecological modeling that will have far reaching impact in the area of Functional and Ecological modeling. The Integrated Learning Paradigm (ILP) is shown to have certain advantages over SVR, Linear Regression, and Random Forest, as represented in various diagrams in this study. As depicted in Figure 2, the residual plots for each model indicate prediction error for the marketing models where residuals for the Integrated Learning Paradigm (ILP) are smaller since they are tightly packed around the zero line.

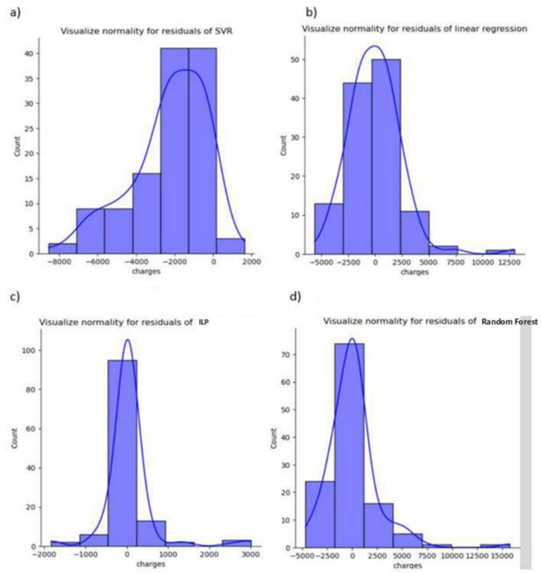

Fig. 3. Visualization of normality for SVR (a), Linear Regression (b), Random Forest (c) and ILP (d).

The level of prediction error in this case was higher than that of the other models, which is a significant advantage for the ILP model. This is in stark contrast to the models, albeit widened, that have alread exhibited a residual spread in the other models. Figure 3 significantly complements this performance in the form of the gross visual aid of residual normality which shows close approximation between the ILP residuals and a normal distribution indicating that the predictions were unbiased and well balanced. In contrast, the deviation from normality in the other models indicated problems with outliers and complex variance. The overlapped histogram in figure 4 depicted a more pronounced peak about zero in ILP's error distribution indicating that there were ILP's greater number of accurate errors more so than in the model where broader error spreads were recorded. These excessive errors did not overshadow the reality that maany of the predictions were accurate over longer periods of time. This is also nettled in detail in table 2 since these findings are further supplemented with ILP being the best model for metrics such as Mean Absolute Error (MAE), RMSE, and R2 thereby further complementing its utility and flexibly.

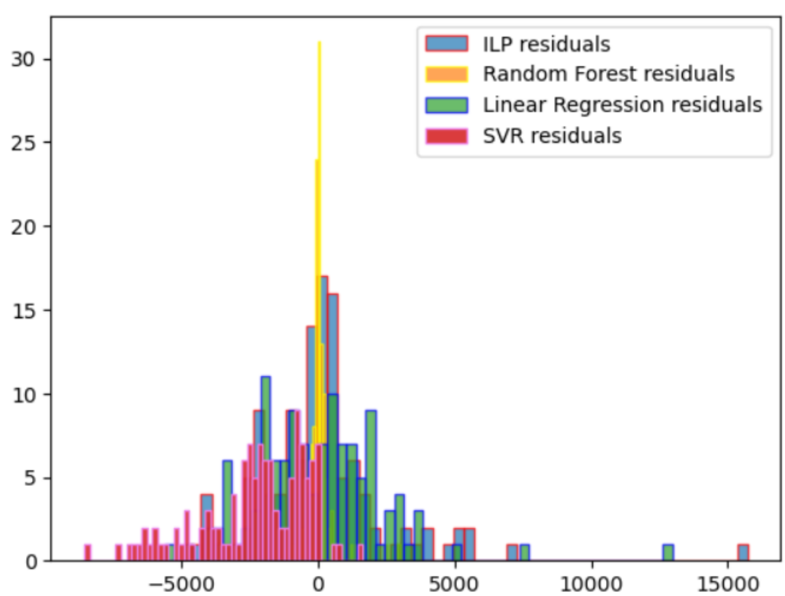

Fig. 4. Overlapped histogram plot for residuals of four algorithms.

TABLE II. COOPERATIVE STUDY OF THE RESULTS

Algorithms	**MAE**	**RMSE**	**R2 Score**
Linear Regression	1803.88	2439.54	0.96
Random Forest Regressor	2344.63	3027.76	0.93
SVR (kernel= ‘linear’)	1608.49	2522.66	0.95
ILP	278.89	580.65	0.99

Once more, the Q-Q plot serves the purpose of illuminating the extent of ILP’s normality, which in turn also strengthens the argument that such model can be predictively valid and consistent across different ecological data. Finally, the histogram of the R^2 scores provided in Figure 6 exhibits the supremacy of ILP practically its theoretical maximal R^2 score stressing the fact that its efficiency in capturing data variability is unprecedented. These visualizations taken together help to further substantiate that ILP is truly a robust, accurate and new model in the context of carrying out ecological forecasts which constitutes new frontiers in the application of ensemble for complex environmental data sets.

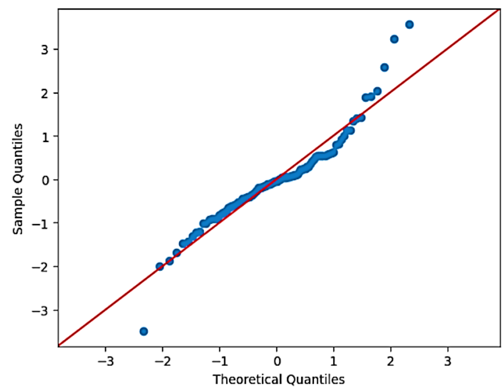

Fig. 5.QQ plot for ILP Algorithm.

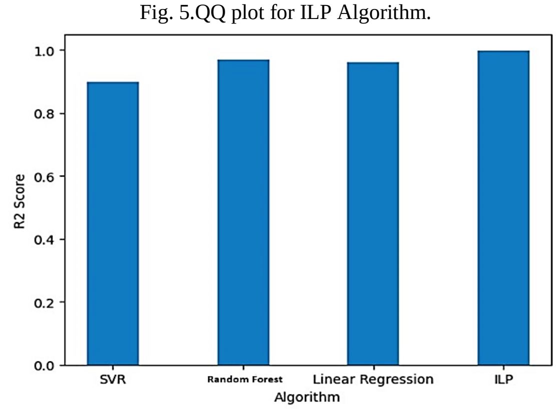

Fig. 6. Histogram Representation of R2 score of four algorithms.

5. CONCLUSION

This research concludes by emphasizing the Integrated Learning Paradigm (ILP) as an innovative tool for enhancing predictive models in the context of ecological datasets. The application of ILP, which permitted the integration of Support Vector Regression (SVR), Linear Regression, and Random Forest into a single system, resulted in significant improvements in model performance, predictive accuracy, and errors when compared to traditional models built with single methods. With an R^2 of 0.9917, ILP was able to account for almost all variance in the dependent variable and was better than each model examined alone. This score indicates that ILP is capable of providing a close approximation of how ecological parameters interact, making the model reliable and accurate enough for ecological effect analysis. Further support is provided by the model’s Root Mean Square Error (RMSE) value of 688.2, which also suggests a dramatic improvement of predictive errors when compared to single algorithms. Such an accuracy level makes ILP to be a paradigm shift in ecological analysis where biometric assessments often involve the analysis of complex, noisier, and high-dimensional data. And indeed, the visual analyses offer further evidence for the success of ILP, showing residual plots where errors are narrowly concentrated about zero predicting little deviation from the target value and hence a finer error distribution.

The normality tests, for instance, the Q-Q plots, show that the ILPs residuals are normally distributed indicating that the model is symmetric and unbiased in its prediction across the different ranges of the datasets. These visual insights in addition to the overlapped histogram of residuals also provide more evidence of ILP’s accuracy bearing in mind that the model has a tendency of clustering her right at the center of the error which is zero, a thing that other models are seldom able to do. ILP cause such narrow error distribution, is of great importance in ecological situations where the effectiveness of prediction would determine the adequate conservation measures and policies to be formulated. The balanced residual behavior draws attention not only to the predictive validity of the model but also to the fact that the model performs well in virtually all types of ecosystems, reflecting once again the versatility of ILP’s integrated framework. The cross-overs carried out in this study in terms of MAE, RMSE and R2 were also effects confirming ILP’s functional efficiency. Although SVR, Linear Regression and Random Forest were each able to perform better in specific areas, ILP was able to harness the best across the broadest range of parameters. The consistent performance on a broad range of evaluation metrics further demonstrates ILP's flexibility in modeling ecological data of varying shapes and complexities.

The minimal MAE especially shows ILP’s ability to produce predictions that have small errors with respect to the average value of the outputs. As a result, virtually every output generated by the method retains the same value for a particular form of the ILP model. This flexibility, therefore, makes ILP a very competent structure for different uses in the ecological biogeography and other respectable disciplines within the sciences where accuracy of prediction is of the essence. To sum up, the present research brings out ILP as a new approach that revolutionizes the setting of the norms of predictive models in the analysis of ecological data. In addition, integrating several regression models into a single one allows ILP to exploit the advantages of the models and combine them in such a revolutionary way to form a new model with better accuracy and reliability. What is further, the success of application of ILP will also have far reaching consequences in that it is possible to apply it in places other than pure research. With the incorporation of this model into conservation approaches, it will greatly increase the role and significance of biodiversity resources. In an era of increasing complexities in ecological data that are critical for sustainable approaches, models such as network add-in ILP will be needed for generating reliable evaluations and effective conservation measures. Future research of ILP may cover its application for the various types of data and different situations in the environment where it may set reference standards in the expanding sphere of ecological forecasting models. This research of course calls for the balancing shift focusing on the development of ensemble models which are useful and at the same time effective for improving prediction analysis tools in the field of environmental science.”

Funding

The authors did not receive support from any organization for the submitted work.

Conflicts of Interest

To their knowledge, the manuscript’s authors, “A Novel Integrated Learning Paradigm for Superior Predictive Modeling in Ecological Data Analysis, “declare no conflict of interest in the present work.

Data Availability Statement

All the data are collected from the survey form.

Acknowledgment

The author extends gratitude to the institution for fostering a collaborative atmosphere that enhanced the quality of this research.

References

[1] Z. Hu and H. Zhang, “Non-selective language activation in L2 lexical inference and text comprehension: Comparing skilled and less-skilled readers,” Heliyon, vol. 9, no. 1, p. e12818, 2023, doi: 10.1016/j.heliyon.2023.e12818.

[2] S. El khediri et al., “Integration of artificial intelligence (AI) with sensor networks: Trends, challenges, and future directions,” J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 1, p. 101892, 2024, doi: 10.1016/j.jksuci.2023.101892.

[3] A. Subeesh and C. R. Mehta, “Automation and digitization of agriculture using artificial intelligence and internet of things,” Artif. Intell. Agric., vol. 5, pp. 278–291, 2021, doi: 10.1016/j.aiia.2021.11.004.

[4] A. K. Singh, J. Patra, M. Chakraborty, and S. Gupta, “Prediction of Indian government stakeholder oil stock prices using hyper parameterized LSTM models,” 2022 Int. Conf. Intell. Controll. Comput. Smart Power, ICICCSP 2022, pp. 1–6, 2022, doi: 10.1109/ICICCSP53532.2022.9862425.

[5] M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, “Detection of Bots in Social Media: A Systematic Review,” Inf. Process. Manag., vol. 57, no. 4, p. 102250, 2020, doi: 10.1016/j.ipm.2020.102250.

[6] L. Breiman, “Random Forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[7] A. Li, H. Xu, C. Feng, H. Yang, and S. Xu, “Incomplete multi-view clustering via local and global bagging of anchor graphs,” Expert Syst. Appl., vol. 248, no. January, 2024, doi: 10.1016/j.eswa.2024.123402.

[8] I. Sengupta, C. Koner, N. K. Bhattacherjee, and S. Gupta, “Automated Student Merit Prediction using Machine Learning,” Proc. - 2022 IEEE World Conf. Appl. Intell. Comput. AIC 2022, pp. 556–560, 2022, doi: 10.1109/AIC55036.2022.9848976.

[9] P. Ratta, Abdullah, and S. Sharma, “A blockchain-machine learning ecosystem for IoT-Based remote health monitoring of diabetic patients,” Healthc. Anal., vol. 5, no. April, p. 100338, 2024, doi: 10.1016/j.health.2024.100338.

[10] F. Chen and G. Fang, “Harnessing digital twin and IoT for real-time monitoring, diagnostics, and error correction in domestic solar energy storage,” Energy Reports, vol. 11, no. August 2023, pp. 3614–3623, 2024, doi: 10.1016/j.egyr.2024.03.024.

[11] M. Awad and R. Khanna, “Efficient learning machines: Theories, concepts, and applications for engineers and system designers,” Effic. Learn. Mach. Theor. Concepts, Appl. Eng. Syst. Des., no. January, pp. 1–248, 2015, doi: 10.1007/978-1-4302-5990-9.

[12] J. M. Durden, B. Hosking, B. J. Bett, D. Cline, and H. A. Ruhl, “Automated classification of fauna in seabed photographs: The impact of training and validation dataset size, with considerations for the class imbalance,” Prog. Oceanogr., vol. 196, no. May, p. 102612, 2021, doi: 10.1016/j.pocean.2021.102612.

[13] S. Luthra, S. Kumar, D. Garg, and A. Haleem, “Barriers to renewable/sustainable energy technologies adoption: Indian perspective,” Renewable and Sustainable Energy Reviews, vol. 41. Elsevier Ltd, pp. 762–776, 2015. doi: 10.1016/j.rser.2014.08.077.

[14] V. Tiwari, R. C. Joshi, and M. K. Dutta, “Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images,” Ecol. Inform., vol. 63, no. March, p. 101289, 2021, doi: 10.1016/j.ecoinf.2021.101289.

[15] B. Mondal and S. Gupta, “Execution Survey and State of the Art of Different ML-Based Ensemble Classifiers Approach Contextual Analysis of Spam Remark Location,” in Proceedings of Third International Conference on Computing, Communications, and Cyber-Security. Lecture Notes in Networks and Systems, vol 421., Vol 421., M. (eds) Singh, P.K., Wierzchoń, S.T., Tanwar, S., Rodrigues, J.J.P.C., Ganzha, Ed. Springer, Singapore, 2023, pp. 311–323. doi: 10.1007/978-981-19-1142-2_24.

[16] S. Bhattacharyya, A. Misra, and K. K. Sarma, “A BCH code assisted modified NCO based LSPF-DPLL topology for Nakagami-m, Rayleigh and Rician fading channels,” Digit. Commun. Networks, vol. 5, no. 2, pp. 102–110, 2019, doi: 10.1016/j.dcan.2017.10.001.

[17] B. Padmaja, V. V. R. Prasad, and K. V. N. Sunitha, “A novel random split point procedure using extremely randomized (Extra) trees ensemble method for human activity recognition,” EAI Endorsed Trans. Pervasive Heal. Technol., vol. 6, no. 22, pp. 1–10, 2020, doi: 10.4108/eai.28-5-2020.164824.

[18] W. Chen, P. Meer, B. Georgescu, W. He, L. A. Goodell, and D. J. Foran, “Feature Extraction for Image Mining,” Comput Methods Programs Biomed, vol. 79, no. 1, pp. 59–72, 2005, doi: 10.1016/j.cmpb.2005.03.006.

[19] C. Peláez-Rodríguez et al., “Deep learning ensembles for accurate fog-related low-visibility events forecasting,” Neurocomputing, vol. 549, 2023, doi: 10.1016/j.neucom.2023.126435.

[20] B. Mondal, A. Banerjee, and S. Gupta, “XSS Filter detection using Trust Region Policy Optimization,” in 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), Jan. 2023, pp. 1–4. doi: 10.1109/ICAISC56366.2023.10085076.

[21] E. Roberson, “Phase Diagram,” Callaloo, vol. 29, no. 1, pp. 20–20, 2006, doi: 10.1353/cal.2006.0067.

[22] D. Swami and D. Parthasarathy, “Dynamics of exposure, sensitivity, adaptive capacity and agricultural vulnerability at district scale for Maharashtra, India,” Ecol. Indic., vol. 121, p. 107206, 2021, doi: 10.1016/j.ecolind.2020.107206.

[23] J. de J. Rugeles Uribe, E. P. Guillen, and L. S. Cardoso, “A technical review of wireless security for the internet of things: Software defined radio perspective,” J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.04.003.