In Table [1], classification error rates are presented for a variety of machine learning models applied to different classification datasets. Each row in the table corresponds to a specific dataset, while the columns represent individual methods: ADAM, BFGS, GENETIC, RBF, NEAT, PRUNE, NNC, and PROPOSED. The values indicate error percentages, meaning that lower values correspond to better model performance on each dataset. The final row shows the average error rate for each model, serving as a general indicator of overall performance across all datasets. Based on the analysis of the average errors, it becomes evident that the PROPOSED method achieves the lowest average error rate, with a value of 19.63%. This suggests that it generally outperforms the other methods. It is followed by the NNC model with an average error of 24.79%, which also demonstrates a significantly lower error compared to traditional approaches such as ADAM, BFGS, and GENETIC, whose average error rates are 36.45%, 35.71%, and 28.25% respectively. The PRUNE method also performs relatively well, with a mean error of 27.94%. On an individual dataset level, the PROPOSED method achieves the best performance (i.e., the lowest error) in a considerable number of cases, such as in the CIRCULAR, DERMATOLOGY, SEGMENT, Z\_F\_S, ZO\_NF\_S, ZONF\_S, and ZOO datasets, where it records the smallest error among all methods. Furthermore, in many of these cases, the performance gap between the PROPOSED method and the others is quite significant, indicating the method's stability and reliability across various data conditions and structures. Some models, including GENETIC, RBF, and NEAT, tend to show relatively high errors in several datasets, which may be due to issues such as overfitting, poor adaptation to non-linear relationships, or generally weaker generalization capabilities. In contrast, the NNC and PRUNE models demonstrate more consistent behavior, while the PROPOSED method maintains not only the lowest overall error but also reliable performance across a wide range of problem types. In summary, the statistical analysis of classification error rates confirms the superiority of the PROPOSED method over the others, both in terms of average performance and the number of datasets in which it excels. This conclusion is further supported by the observation that the PROPOSED method achieves the best results in the majority of datasets, often with significantly lower error rates. Such superiority may be attributed to better adaptability to data characteristics, effective avoidance of overfitting, and, more broadly, a more flexible or advanced algorithmic architecture.

Table [2] presents the performance of various machine learning methods on regression datasets. In this table, columns represent different algorithms, and rows correspond to datasets. The numerical values shown are absolute errors, indicating the magnitude of deviation from the actual values. Therefore, smaller values signify higher prediction accuracy for the corresponding model. The last row reports the average error for each method across all datasets, offering a general measure of overall performance. According to the overall results, the PROPOSED method exhibits the lowest average error value at 4.83, indicating high accuracy and better overall behavior compared to the other approaches. The second-best performing model is NNC, with an average error of 6.29, which also stands out from the traditional methods. On the other hand, ADAM and BFGS show significantly higher error rates, at 22.46 and 30.29 respectively, suggesting that these methods may not adapt well to the specific characteristics of the regression problems evaluated. At the individual dataset level, the PROPOSED method achieves notably low error values across multiple datasets, including AIRFOIL, CONCRETE, LASER, PL, PLASTIC, and STOCK, outperforming other algorithms by a considerable margin. Its consistent performance across such diverse problems suggests that it is a flexible and reliable approach. Furthermore, the fact that it also performs strongly on more complex datasets with high variability in error—such as AUTO and BASEBALL—strengthens the impression that the method adapts effectively to varying data structures. By comparison, algorithms such as GENETIC and RBF exhibit less stable behavior, showing good performance in some datasets but poor results in others, resulting in a higher overall average error. The PRUNE method, although not a traditional algorithm, shows moderate performance overall, while NEAT does not appear to stand out in any particular dataset and also maintains a relatively high average error. In conclusion, the analysis indicates that the PROPOSED method clearly excels in predictive accuracy, both on average and across a large number of individual datasets. Its ability to minimize error across different types of problems makes it a particularly promising option for regression tasks involving heterogeneous data.

Table [3] presents the error rates of the proposed machine learning model on various classification datasets, considering four different values of the parameter Iw (initialization factor): 2, 3, 5, and 10. The recorded values correspond to error percentages for each dataset, while the last row of the table includes the average error rate for each parameter value. Analyzing the data, it is observed that the value Iw=10 exhibits the lowest average error rate (19.63%), followed by Iw=5 (19.89%). The values Iw=2 and Iw=3 have slightly higher averages, 20.32% and 20.33% respectively. The difference between the averages is relatively small, a fact suggesting that the parameter Iw does not dramatically affect the model's performance; however, the gradual decrease in average error with increasing parameter value may indicate a trend of improvement.

In individual datasets, small variations are observed depending on the setting. In some cases, such as SEGMENT and CIRCULAR, increasing the parameter value leads to noticeably better results. For example, in SEGMENT the error rate decreases from 39.10% for Iw=2 to only 9.59% for Iw=10. A similar improvement is observed in CIRCULAR, where the error decreases from 14.71% to 4.22%. Conversely, in other datasets the variation in values is smaller or negligible, and in some cases, such as ECOLI and CLEVELAND, higher Iw values lead to slightly increased error. Overall, the statistical analysis shows that although no statistically significant differences are observed between the different parameter values, in accordance with the p-values from previous analyses, there is nevertheless an indication that higher values of Iw, such as 10, are associated with slightly improved average performance and better results in certain datasets. This trend may be interpreted as an indication that a higher initialization factor might allow the model to start from more favourable learning conditions, particularly in datasets with greater complexity. However, because the variation is not systematic across all datasets, the selection of the Iw value should be done carefully and in relation to the characteristics of each specific problem.

In Table [4], a general trend of decreasing average error is observed as the value of the initialization factor Iw increases. The average drops from 6.08 (for Iw=2) to 5.48 (Iw=3), 5.24 (Iw=5), and finally 4.83 (Iw=10). This sequential decrease suggests that higher values of Iw tend to improve the model's overall performance. However, the effect is not uniform across all datasets. In some cases, the improvement is striking: in AUTO the error decreases from 17.16% (Iw=2) to 11.73% (Iw=10), in HOUSING it reduces from 27.19% to 15.96%, and in FRIEDMAN the most noticeable improvement is recorded from 6.49% to 1.25%. Additionally, in STOCK a significant drop from 8.79% to 3.96% is observed. Conversely, in some datasets performance deteriorates with increasing Iw: in BASEBALL the error increases from 59.05% (Iw=2) to 60.42% (Iw=10) and in LW from 0.11 to 0.32. In other datasets, such as AIRFOIL, LASER, and PL, differences are minimal and practically negligible, with values remaining very close for all Iw parameters. For example, in AIRFOIL all values are around 0.002, while in PL the difference between values is merely 0.001. This heterogeneity in the response of different datasets underscores that the optimal value of Iw depends significantly on the specific characteristics of each problem. Despite the general improving trend with higher Iw values, notable exceptions like BASEBALL and LW confirm that there is no universal optimal setting suitable for all regression problems.

To determine the significance levels of the experimental results presented in the classification dataset tables, statistical analyses were conducted using scripts written in the R programming language. These analyses were based on the critical parameter "p", which is used to assess the statistical significance of performance differences between models. As shown in Figure [stat1.pnp], the differences in performance between the PROPOSED model and all other models namely ADAM, BFGS, GENETIC, RBF, NEAT, and PRUNE are extremely statistically significant with p < 0.0001. This indicates, with a high level of confidence, that the PROPOSED model outperforms the rest in classification accuracy. Even the comparison with NNC, which is the model with the closest average performance, showed a statistically significant difference with p < 0.05. This confirms that the superiority of the PROPOSED model is not due to random variation but is statistically sound and consistent. Therefore, the PROPOSED model can be confidently considered the best choice among the evaluated models for classification tasks, based on the experimental data and corresponding statistical analysis.

From the analysis of the results presented in Figure [stat2.pnp], it is evident that the performance difference between the PROPOSED model and BFGS is extremely significant (p < 0.0001), clearly indicating the superiority of the PROPOSED model. Similarly, the comparisons with GENETIC and NEAT show very high statistical significance (p < 0.001), confirming that the PROPOSED model achieves clearly better results. The difference with NNC, though smaller, remains significant (p < 0.01), showing that even in comparison with one of the best-performing alternative models, the PROPOSED model still outperforms. The differences with ADAM, RBF, and PRUNE are statistically significant at the p < 0.05 level, suggesting a noteworthy advantage of the PROPOSED model in these cases as well, albeit with a lower confidence level. Overall, the statistical analysis of the regression dataset results confirms the overall superiority of the PROPOSED model, not only in terms of average prediction accuracy but also in the consistency of its performance compared to the alternative approaches.

Figure [stat3.pnp] presents the significance levels for the comparison of different values of the Iw (Initial weights) parameter in classification datasets. The comparisons include the pairs Iw=2 vs Iw=3, Iw=3 vs Iw=5, and Iw=5 vs Iw=10. In all cases, the p-values are greater than 0.05, indicating that the differences between the respective settings are not statistically significant. This implies that varying the Iw parameter across these specific values does not substantially affect the model's performance in classification tasks, and thus, no significant changes in outcomes are observed due to this parameter.

In Figure [stat4.pnp], the statistical evaluation focuses on how different initial weight settings (Iw) affect performance in regression tasks. The comparisons between the values Iw=2, Iw=3, Iw=5, and Iw=10 revealed no significant variations, as all corresponding p-values were found to be greater than 0.05. This outcome suggests that altering the Iw parameter within this range does not lead to measurable differences in the models’ predictive behavior. The results imply that model accuracy remains stable regardless of these specific Iw configurations in regression scenarios.