Journal of Quality Engineering and Production Optimization

Volume 10, Issue 1, Winter & Spring 2025

http://jqepo.shahed.ac.ir

Research Paper

DOI: 10.22070/jqepo.2025.20710.1304

A Hybrid Model for Customer Churn Prediction: An Optimized Combination of Multilayer Perceptron and Atomic Orbital Search

Leila Zolqadr [1], Majid Shakhsi-Niaei [1*], Azam Safari [2]

[1] *Department of Industrial Engineering, Yazd University, Yazd, Iran*

[2] *Department of Information Technology, Holy Shrine Fatemeh Masoumeh (SA), Qom, Iran*

*** Corresponding Author:** *Majid Shakhsi-Niaei (Email: m.niaei@yazd.ac.ir)*

Abstract –*Customer churn is a formidable and persistent challenge for the telecommunications sector, as it significantly erodes profitability and impedes sustainable competitive advantage. Conventional machine learning methods often exhibit limitations in capturing non-linear intricacies and effectively managing voluminous, high-dimensional datasets. To address these issues, this study presents a novel hybrid predictive framework for enhanced customer churn prediction. Our proposed methodology integrates an optimized Multilayer Perceptron (MLP) with the Atomic Orbital Search (AOS) metaheuristic algorithm. Crucially, AOS is systematically deployed to fine-tune the critical connection weights and biases of the MLP architecture, mitigating issues such as premature convergence and suboptimal parameter selection that are common in standalone neural networks. Empirical validation, conducted using the comprehensive Teldata dataset (7043 instances, 20 features) in MATLAB, unequivocally demonstrates the hybrid approach's superior efficacy over established conventional methodologies. Specifically, the hybrid MLP-AOS model achieved an impressive predictive accuracy of 97.8% on the test subset, a notable improvement over the 79% achieved by Support Vector Machines (SVMs) and the 75% achieved by a conventional MLP. These compelling findings underscore the proposed approach's ability to predict customer churn with heightened precision, providing telecommunications management with an analytical tool for identifying pivotal influencing factors and formulating effective retention strategies.*

Keywords*– Customer Churn Prediction, Multilayer Perceptron (MLP), Atomic Orbital Search (AOS), Evolutionary Optimization, Hybrid Model, Support Vector Machines (SVMs).*

I. INTRODUCTION

The telecommunications sector is one of the most dynamic and rapidly evolving industries, driven primarily by its adaptation to modern lifestyles, which are increasingly reliant on pervasive internet connectivity and advanced communication technologies. Despite this impressive growth, telecommunication companies are profoundly impacted by two critical challenges: escalating customer churn, and increasing market saturation. In this highly competitive landscape, customer churn-defined as the tendency of subscribers to switch from one service provider

Manuscript Received: 9-Sep-2024 & Revised:5-Dec-2024 & Accepted: 5-Feb-2025
ISSN: 2423-3781

to a competitor, represents a significant drain on profitability and long-term viability. The fundamental importance of customer retention is underscored by the fact that existing customers constitute the most vital source of revenue for these organizations (Noe et al., 2017). Consequently, a sustained high churn rate can erode a company's competitive advantage and, if left unaddressed, may even jeopardize its survival when the rate of new customer acquisition fails to compensate for departures (Pustokhina et al., 2021).

Customer churn, a critical metric in customer relationship management, is typically categorized into two primary types: voluntary and involuntary churn. Voluntary churn occurs when a customer deliberately chooses to terminate their contractual agreement with a service provider (Hossein Khani et al., 2014). This can manifest as either intentional or accidental. Intentional voluntary churn often stems from customer dissatisfaction, competitive offers, unfavorable pricing, suboptimal customer service, or network issues. Conversely, accidental voluntary churn may arise from a diminished need for the service or relocation outside the service provider's operational area. While accidental churn, which also encompasses issues such as illness, death, fraud, or bad debt, generally accounts for a smaller proportion of voluntary departures, intentional churn is the primary focus of most attrition management strategies (Hossein Khani et al., 2014). In contrast, involuntary churn is initiated by the company itself, often due to a customer's failure to meet contractual obligations, such as financial defaults. Effective churn management, a widely recognized imperative within the telecommunications sector, encompasses all actions undertaken to mitigate customer attrition (Ren et al., 2019). This concept is inherently linked to customer retention, which involves a customer continuing their patronage with a supplier. Customer retention management, in turn, refers to strategies employed by companies to engage and retain their customer base. Given their inverse relationship, successful customer retention paradigms often adhere to two fundamental principles: the maximization of profitability through increased sales and the optimization of costs leading to achieve competitive pricing and sustained customer loyalty (Alsiehemy, 2019).

A significant challenge confronting telecommunications enterprises resides in their capacity to effectively analyze the intricate patterns and behaviors of their customer base through the examination of vast customer-centric datasets. Big data, in this context, refers to extensive and complex data aggregations that exceed the processing capabilities of conventional information technology techniques and applications (Boobier, 2018). The extraction of meaningful insights from these voluminous datasets is paramount, as the identified patterns are instrumental in facilitating robust data classification, accurate predictive modeling, and ultimately, the construction of highly relevant and effective data-driven models (Ever et al., 2019; Huettmann et al., 2018).

To mitigate the detrimental effects of customer churn and support customer retention, employing various machine learning (ML) techniques has become essential (Deng et al., 2019). Recognizing that traditional predictive models are often limited in their capacity to handle the complex, nonlinear relationships within vast customer datasets, this research seeks to advance the current methodology for churn prediction. Our primary goal is to develop a robust and accurate model for predicting customer churn through a sophisticated blend of neural network algorithms and advanced optimization methods. This study specifically aims to accomplish the following objectives:

- To adaptively optimize the vital weights and biases of the MLP architecture by using the network's error rate as the objective function for the AOS optimization algorithm;
- To improve the predictive capability of the MLP through a seamless integration with the AOS algorithm, thus establishing a synergistic and powerful predictive framework;
- To meticulously assess the accuracy of the proposed customer churn prediction model by examining the parameters that were precisely tuned within the network.

The core contribution of this work lies in the strategic combination of the AOS algorithm, a relatively new metaheuristic, to specifically and adaptively fine-tune the complex connection weights and biases of the MLP neural network for enhanced customer churn prediction. This approach uniquely addresses persistent issues in MLP training, such as preventing arbitrary weight initialization and mitigating the model's tendency to get trapped in local optima during the learning process. The final optimized model, a fusion of AOS and MLP, represents a new and effective framework that enhances predictive accuracy and stability across intricate, real-world telecommunications datasets.

II. LITERATURE REVIEW AND RESEARCH GAP

A. Literature Review

The landscape of customer churn prediction has undergone significant evolution, driven by the increasing availability of data and advancements in analytical methodologies aimed at mitigating its detrimental effects on telecommunication profitability. Early efforts primarily focused on understanding foundational churn dynamics and employing conventional statistical or simpler machine learning models.

Initially, studies such as Lee et al. (2001) investigated fundamental factors, such as customer satisfaction and switching costs, and concluded about their impact on churn probability. While providing foundational insights, such early models often had limitations in detection accuracy. Subsequently, Neslin et al. (2006) explored methodological factors influencing churn prediction accuracy using decision trees and ANNs, noting the robustness of such models but also acknowledging their time-consuming nature and sometimes lower accuracy. Han et al. (2009) approached the issue from a qualitative perspective, using structural equation analysis to reveal the interplay of customer emotions and switching barriers on retention, though their qualitative method had limitations in reassessing variable correlations.

As computational capabilities advanced, more sophisticated machine learning algorithms began to emerge. Sato et al. (2010) proposed using Principal Component Analysis (PCA) for churn prediction, demonstrating its superior performance over methods such as Naive Bayesian, SVM, and traditional decision trees on real-world datasets, despite PCA's challenges with overfitting and normalization. Adwan et al. (2014) delved into applying MLP neural networks with backpropagation for churn prediction, identifying the absence of optimal settings for weight selection and neuron adjustment as a key limitation, even while showing low execution time. Hossein Khani et al. (2014) also contributed by identifying factors affecting churn in the insurance industry, proposing a rule-based decision-making technique.

With the rise of big data in the mid-2010s, the focus shifted towards algorithms capable of handling massive and complex datasets. Bi et al. (2016) proposed a semantic-centered reduction clustering method for mitigating churn risk in big data environments, which showed stronger clustering power but faced complexity, overfitting, and issues with missing data. Amin et al. (2017) utilized a rough set approach to classify churn, offering improved detection accuracy over genetic methods. These developments underscored the growing necessity for robust algorithms that could manage the sheer volume and dimensionality of contemporary telecom data.

More recently, research has focused on improving predictive performance and addressing specific challenges, such as data imbalance and model interpretability. Huettmann et al. (2018) highlighted the broader applications of machine learning for analyzing large ecological datasets, providing an overview of its potential. Boobier (2018) discussed the impact of advanced analytics and AI on various sectors, including the need for effective data management. Ever et al. (2019) contributed by comparing different machine learning techniques for prediction problems, further emphasizing the role of big data. Ren et al. (2019) examined factors such as product variety and market structure in relation to churn, providing insights into competitive dynamics.

In the 2020s, a strong emphasis emerged on highly optimized and specialized techniques. Rai et al. (2020) applied decision tree classification for churn prediction, noting challenges related to computational complexity as the number of variables increased. Lalwani et al. (2021) developed a customer churn prediction system using a machine learning approach and demonstrated its effectiveness. Pustokhina et al. (2021) introduced dynamic churn prediction strategies using text analytics and evolutionary optimization algorithms. Ramesh et al. (2022) proposed a hybrid ANN and Random Forest approach for churn prediction, achieving high accuracy with telecom data. Chugh et al. (2022) focused on optimizing deep learning for analyzing unstructured social media data to predict churn, albeit noting high execution times for large datasets.

Recent advancements in the field continue to push the frontiers of both accuracy and robustness, particularly for the significant challenge of customer churn, by embracing new methodologies and application areas. Bhushan et al. (2024) introduced a robust and explainable machine learning model tailored specifically for churn prediction in the telecommunication industry. Their work emphasizes the critical importance of achieving both high predictive accuracy and model interpretability in this sector. In parallel, Maskale et al. (2024) detailed the design and implementation of an application for predicting and analyzing customer churning rates in the banking sector. Their research showcased the use of various machine learning algorithms, including logistic regression, random forests, and gradient boosting, to effectively identify potential churners and formulate retention strategies. A comprehensive review of 212 published articles on churn prediction by Manzoor et al. (2024) found that the profitability aspect of churn models remains under-researched. They advocated the use of profit-based evaluation metrics and recommended adopting ensembles, deep learning, and explainable methods to guide future advancements, thereby highlighting current research gaps.

Looking towards the immediate future, Sagming et al. (2025) introduced Topological Data Analysis (TDA) to uncover hidden structural features, significantly improving XGBoost model performance for churn prediction. Chai et al. (2025) rigorously addressed data disparities and imbalances, showing how techniques like SMOTE can enhance model accuracy by mitigating bias in imbalanced datasets. Similarly, Zhang et al. (2025) focused on data-driven approaches for improved churn prediction through customer segmentation, achieving high accuracy.

Table I. Summary of relevant literature, compared to the present paper

Year	Researchers	Method	Advantages	Disadvantages
2001	Lee et al.	Customer Satisfaction and Switching Cost Analysis	Low execution time	Low detection accuracy
2006	Neslin et al.	Decision Trees and ANN	Variable identification	Time-consuming, and low accuracy
2009	Han et al.	Structural Equation Analysis	Effective emotional impact identification	Failure to reassess variable correlations
2010	Sato et al.	Principal Component Analysis	Dimensionality reduction	Overfitting, and poor normalization
2014	Adwan et al.	Multilayer Perceptron Neural Networks	Low execution time	Suboptimal parameter selection, and neuron adjustment issues
2016	Bi et al.	Semantic-Based Reduction Clustering	Higher power compared to fuzzy methods	Computational complexity
2017	Amin et al.	Rule-Based Decision-Making with Rough Set Theory	Medium accuracy, competitor, and customer insights	Requirement of expert knowledge, law-based constraints
2019	Rai et al.	Decision Tree Classification	Comprehensive outcomes analysis	High computational complexity, imbalance issues
2022	Chugh et al.	Query Analysis with Deep Learning and Optimization	High accuracy	High execution time, heavy hardware requirements
2022	Ramesh et al.	Hybrid Artificial Neural Networks and Random Forests	Increased accuracy, effective solutions, telecom data usage	Computational complexity, parameter tuning
2024	Manzoor et al.	Comprehensive Review on ML Methods for Churn Prediction	Extensive coverage of existing research	Focus on review, not new model development.

Continue Table I. Summary of relevant literature, compared to the present paper

Year	Researchers	Method	Advantages	Disadvantages
2024	Maskale et al.	Machine Learning Application for Churn Prediction	Identifies potential churners, devises effective retention strategies, and develops domain-specific applications	Lack of algorithmic novelty, primary focus on application design rather than in-depth model comparison or theoretical contribution.
2024	Bhushan et al.	Robust and Explainable Machine Learning Model	High predictive accuracy, model interpretability, and robust performance	Computational complexity, potential for extensive hyperparameter tuning, and limited testing on diverse datasets for generalizability.
2025	Chai et al.	Imbalance Correction Techniques and Machine Learning	Improved accuracy, reduced bias, and diverse methods	Computational complexity, data limitations, parameter tuning
2025	Sagming et al.	Topological Data Analysis (TDA) and Machine Learning	Improved accuracy, hidden feature extraction, and reduced parameter tuning	Computational complexity, implementation challenges
2025	Zhang et al.	Data-Driven Customer Segmentation and Predictive Modeling	High accuracy, customer segmentation, and increased profitability	Computational complexity, parameter tuning, and generalizability challenges
2025	This article	New Method of Combining Evolutionary Optimization and Advanced Neural Networks	High accuracy, innovative approach, and adaptable to real-world data	Computational complexity

B. Research Gap and Contributions

Despite significant advancements in customer churn prediction, several persistent challenges and research gaps remain evident in the existing literature. A primary limitation is the common susceptibility of traditional MLPs to suboptimal performance. This often stems from their reliance on arbitrary weight initialization and their tendency to converge to local optima during complex training processes. While various metaheuristic algorithms have been explored for optimization, a comprehensive investigation into the precise application of newer, globally robust metaheuristics, such as the AOS algorithm, for fine-tuning MLP parameters in the context of customer churn remains largely unexplored.

Furthermore, a recurring issue in real-world churn datasets, particularly in telecommunications, is the inherent imbalance in class distribution. Although some studies have addressed this using balancing techniques, there remains a need for models that demonstrate robust performance across both majority and minority classes without resorting to complex resampling strategies that might alter the original data distribution. Additionally, many existing models, while achieving high accuracy, often lack interpretability, making it challenging for business practitioners to understand the key factors driving churn and to formulate targeted retention strategies. The integration of advanced optimization with interpretable model components, or methods for feature importance analysis, represents a crucial yet under-explored avenue.

In response to these identified gaps, this study offers several significant and unique contributions to the field of customer churn prediction:

- Development of a New Hybrid Model: A new hybrid model is proposed that synergistically integrates the Atomic Orbital Search (AOS) algorithm with a Multilayer Perceptron (MLP) neural network. This specific fusion is designed to adaptively optimize the intricate connection weights and biases of the MLP, thereby mitigating issues of random initialization and local optima prevalent in standalone MLP training.

- Enhanced Predictive Performance and Robustness: Our methodology thoroughly assesses the model’s performance using a full range of classification metrics (Accuracy, Precision, Recall, F1-Score) on a real-world telecommunications dataset with inherent class imbalance. The empirical findings indicate superior and robust predictive accuracy, with particularly strong performance on the minority (churn) class, thereby confirming the model’s high effectiveness in challenging real-world contexts.
- Optimization for Complex Parameter Spaces: This research highlights the efficacy of AOS in navigating the high-dimensional parameter space of MLP weights, offering a refined approach to model training that aims for globally optimal solutions.
- Practical Managerial Insights: The study emphasizes the practical applicability of the proposed model by providing a framework that can assist telecommunications managers in identifying pivotal influencing factors of churn, thereby enabling more targeted and effective customer retention paradigms.
- Comparative Analysis with Baselines: A thorough comparison with conventional machine learning algorithms (Simple MLP and SVM) is provided, clearly demonstrating the performance advantages of the proposed optimization approach.

III. RESEARCH METHOD

A. Atomic Orbital Search (AOS)

Traditional regression methods and their extended variants have been widely applied to analyze relationships among variables influencing customer churn rate prediction. However, these methods exhibit inherent challenges, including an inability to effectively model nonlinear data relationships, sensitivity to internal data structure, and the absence of a backpropagation mechanism (Rani et al., 2021). In contrast, neural network-based approaches demonstrate a superior capacity to accurately analyze data relationships by considering both internal and external data structures. It is noteworthy that the simplest form of a neural network, lacking a hidden layer, essentially functions as a regression method. Therefore, evolving regression techniques into more sophisticated neural networks and subsequently enhancing these networks can significantly improve the prediction of factors affecting the customer churn rates (Lalwani et al., 2021). Nevertheless, standalone neural networks often exhibit limitations, such as insufficient predictive power and susceptibility to premature convergence, primarily due to the arbitrary initialization of connection weights. The selection of MLPs as the core predictive model is driven by their proven ability to effectively model complex non-linear relationships and handle large, high-dimensional datasets, which are characteristic of customer churn prediction problems. However, traditional MLPs often face challenges such as random weight initialization leading to suboptimal performance, and susceptibility to local optima during the training process (Goodfellow et al., 2016; Nielsen, 2015). To mitigate these inherent limitations, we employed the AOS algorithm for its robust optimization capabilities. While widely used metaheuristic algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have demonstrated success in various optimization tasks, AOS, as a more recently developed approach, offers a unique search mechanism inspired by quantum mechanics. This mechanism provides an exceptional balance between exploration (searching new areas of the solution space) and exploitation (refining existing good solutions), making it particularly suitable for fine-tuning the complex, high-dimensional parameter space of MLP weights and biases. This approach aims to overcome the issues of local optima and arbitrary initialization that can hinder other metaheuristics in similar complex optimization challenges (Azizi, 2020).

To address these limitations, metaheuristic optimization methods are highly instrumental. These algorithms are designed to generate a diverse set of candidate solutions for a given problem, subsequently exploring and exploiting these solutions to identify the optimal one, thereby facilitating overall problem optimization. Consequently, this study employs the AOS method to enhance the predictive capabilities of neural networks (Azizi, 2022).

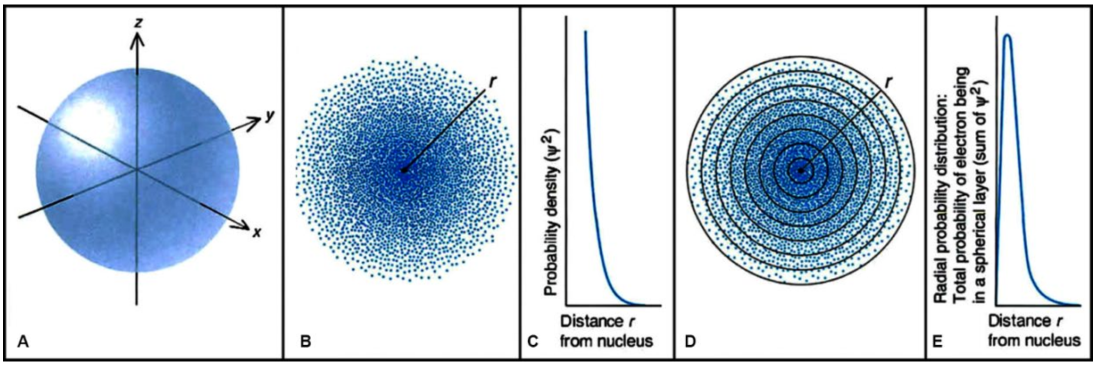


Fig. 1. Configuration of the electron density of an atom (Adapted from Azizi, 2020).

The theoretical underpinning of the AOS algorithm is rooted in classical atomic models. Atomic orbitals are conceptualized as regions surrounding an atom's nucleus where the probability of locating an electron is highest (Fig. 1A). Electrons, characterized as charge clusters, exhibit rapid positional changes over time (Fig. 1B). Although an electron's exact position around the nucleus is indeterminable at any given moment, its probability of presence can be precisely calculated through a probability density diagram (Fig. 1C). The space around the nucleus can be envisioned as concentric layers, which allows for the estimation of the overall probability of finding an electron at various distances (Fig. 1D). As demonstrated in Fig. 1E, the radial probability distribution suggests a higher likelihood of locating an electron in the second layer than in the first. This model is based on the assumption that the atom's electrons are at their ground state energy level. The energy level of an electron is defined by the principal quantum number '*n*', with higher '*n*' values correlating to orbitals of larger radii and greater energy (Azizi, 2020).

To extract an electron from its atomic shell, a specific quantum of energy, termed binding energy, is required. In the context of the AOS algorithm, the electron cloud encircling the nucleus metaphorically represents the problem's search space, wherein each electron corresponds to a candidate solution (X_i). The positional vectors, denoted as $x(i,j)$, are employed to delineate the precise locations of these candidate solutions within the search domain. Here, '*m*' signifies the total number of solutions (electrons) within this search space, and '*d*' represents the problem's dimensionality.

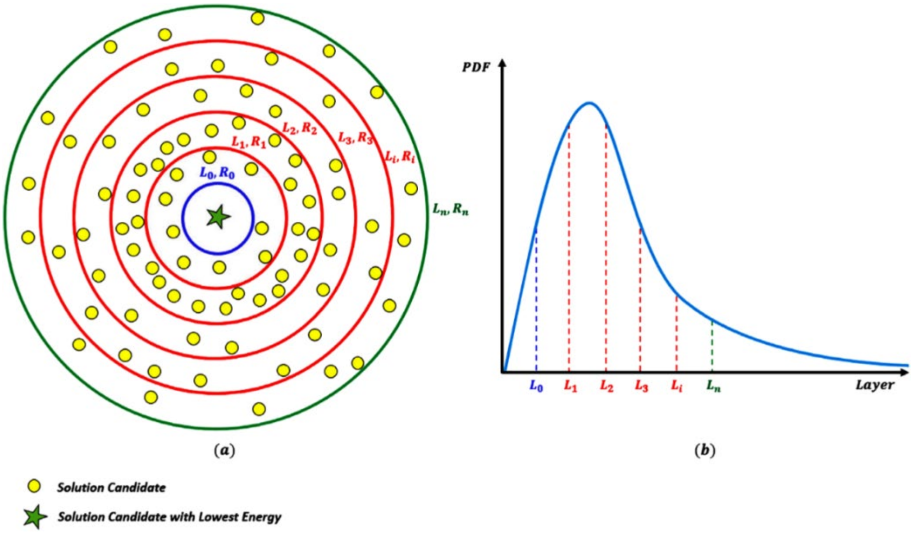


Fig. 2. Electron density configuration of an atom (Adapted from Azizi, 2020).

The process for determining the positions of candidate solutions (electrons) across the hypothetical layers is schematically illustrated in Fig. 2. This is accomplished by utilizing a probability density function derived from the normal Gaussian distribution. The mathematical formulations for the position vectors (X^k) and their corresponding objective function values (E^k) of candidate solutions within these layers are also provided. Specifically, the position of the i-th solution in the k-th hypothetical layer is denoted as X_i^k, while 'p' indicates the total number of solutions within that layer and 'd' represents the problem's dimension. The variable 'n' refers to the total number of hypothetical layers in the model. The objective function value for the i-th solution is represented by E_i^k. Furthermore, the binding state and binding energy of the k-th layer are denoted by BS^k and BE^k, respectively. The variables X_i^k and E_i^k symbolize the position and fitness of the i-th solution in the k-th layer. Ultimately, the total count of all solutions within the expansive search space is quantified by the variable 'm'. These definitions are underpinned by the following precise mathematical equations:

$$X^k = \begin{bmatrix} X_1^k \\ X_2^k \\ \cdot \\ \cdot \\ X_i^k \\ \cdot \\ \cdot \\ X_p^n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 \dots & x_1^j & \dots & x_1^d \\ x_2^1 & x_2^2 \dots & x_2^j & \dots & x_2^d \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_i^1 & x_i^2 \dots & x_i^j & \dots & x_i^d \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_p^1 & x_p^2 \dots & x_p^j & \dots & x_p^d \end{bmatrix}, \quad \begin{cases} i = 1,2,..,p \\ j = 1,2,..,d \\ k = 1,2,..,n \end{cases} \tag{1}$$

$$E^k = \begin{bmatrix} E_1^k \\ E_2^k \\ \cdot \\ \cdot \\ E_i^k \\ \cdot \\ \cdot \\ E_p^n \end{bmatrix}, \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \tag{2}$$

$$\begin{aligned} BS^k &= \frac{\sum_{i=1}^{p} X_i^k}{p}, \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \\ BE^k &= \frac{\sum_{i=1}^{p} E_i^k}{p}, \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \end{aligned} \tag{3}$$

Within this model, the energy level (E_i^k) of each solution is conceptually compared to the layer's binding energy to govern the emission or absorption of photons. Photon emission is considered when a candidate solution's energy level (X_i^k) is equal to or greater than the binding energy of its layer (BE^k). Conversely, if the energy level falls below the layer's binding energy, photon absorption is taken into account (Azizi, 2022). The position updates for candidate solutions are then defined by the following mathematical equations:

$$X_{i+1}^k = X_i^k + \frac{\alpha_i * (\beta_i * LE - \gamma_i * BS)}{k}, \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \tag{4}$$

$$X_{i+1}^k = X_i^k + \alpha_i * (\beta_i * LE^k - \gamma_i * BS^k), \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \tag{5}$$

The process for updating the position of each solution is defined in the following manner, based on the aforementioned effects. Specifically, X_i^k and X_{i+1}^k correspond to the current and subsequent positions of a solution at the i-th iteration and within the k-th layer. A random vector, r_i, drawn from a uniform distribution between 0 and 1, is also utilized in this process (Azizi, 2020).

$$X_{i+1}^{k} = X_{i}^{k} + r_i, \quad \begin{cases} i = 1,2,..,p \\ k = 1,2,..,n \end{cases} \tag{6}$$

B. Optimization of Multilayer Perceptron Neural Network Using Atomic Orbital Search

A salient approach to enhancing the efficacy of MLP neural networks involves the precise optimization of the connection weights interlinking the neurons across their various layers. It is critical to note that simultaneously optimizing both the network structure and weights within an MLP significantly increases the number of tunable parameters, thereby leading to a substantial increase in computational complexity and execution time. Given this challenge, the present study exclusively concentrates on optimizing the weights associated with the connections between the input and hidden layer neurons, as well as between the hidden output layer neurons, specifically within an MLP neural network trained via backpropagation (Goodfellow et al., 2016; Nielsen, 2015).

The following outlines the steps of the optimized MLP method for predicting customer churn rates:

1. Create a neural network with Bayesian backpropagation training on the desired telecommunications database.
2. Extract the number of weights and biases in the network.
3. Optimize the connection weights using the AOS method. The stages for the optimization are described as follows:
   - Initialize the initial positions of the solutions (X_i) in the problem's search space.
   - Evaluate the fitness of each solution's position (E_i).
   - Determine the connection state (BS) and binding energy (BE) of the atom.
   - Identify the solution with the lowest energy level in the atom (LE).
   - Until the convergence condition is met, perform the following steps:
4. Generate and create n hypothetical layers.
5. Sort the problem solutions in ascending or descending order.
6. Distribute the solutions in the hypothetical layers using the probability density function.
7. For each hypothetical layer, execute the following steps:
   - Determine the connection state (BS^k) and binding energy (BE^k) of each layer k.
   - Identify the solution with the lowest energy level in layer k.
8. Determine the connection state (BS^k) and binding energy (BE^k) of each layer k.
   - Identify the solution with the lowest energy level in layer k.
   - For each determined maximum iteration, execute the following steps:
   - Initialize the tuning parameters ϕ, ,α, β, and γ.
   - Calculate the parameter PR.
   - If ($\phi \geq PR$), proceed to step 19; otherwise, go to step 23.
   - If $E_i^k \geq BE^k$Execute step 20; otherwise, go to step 21.
   - $X_{i+1}^{k} = X_i^k + \frac{\alpha_i * (\beta_i * LE - \gamma_i * BS)}{k}$
   - If $E_i^k \leq BE^k$, proceed to step 22; otherwise, return to step 18.
   - $X_{i+1}^{k} = X_i^k + \alpha_i * (\beta_i * LE^k - \gamma_i * BS^k)$
   - If $\phi < PR$Go to step 24.
   - $X_{i+1}^{k} = X_i^k + r_i$
9. Sort the solutions, and if the convergence condition is met, record the best solution and exit; otherwise, return to step 5.
10. Extract the best solution from the AOS; the best solution represents the optimized connection weights in the MLP neural network.

The final step involves utilizing the refined, optimized connection weights to configure the MLP neural network, enabling it to predict customer churn.

C. Customer Churn Rate Prediction Using an Enhanced Neural Network

For effective problem prediction utilizing an MLP neural network, the precise configuration of its input, hidden, and output layers is paramount (Nielsen, 2015).

A key requirement is that the problem-specific data must be appropriately mapped onto the input and output neurons of the network (Goodfellow et al., 2016). The architecture examined in this study consists of a three-layer perceptron neural network, featuring an input layer, a hidden layer, and an output layer. Neurons within the hidden layer receive a combination of signals from the input layer to form an estimation. An activation function is then applied to these aggregated signals, and the resulting values are transmitted to the neurons of the output layer through a series of learned connection weights (Nielsen, 2015; Goodfellow et al., 2016; Géron, 2022).

Given that the primary objective is customer churn rate prediction, the Teldata telecommunications database serves as the empirical foundation for this study. This comprehensive dataset encompasses 7043 customer samples and 20 distinct customer features. In this context, individual customers are treated as data samples, while their attributes are considered data features or independent variables.

Accordingly, the churn outcome is designated as the dependent variable (response). The independent variables are consequently represented by the input layer neurons, with the dependent variable forming the output layer neurons. For this specific implementation, the input layer consists of 19 neurons, which corresponds to the count of customer attributes in the Teldata database, while the output layer contains a single neuron that signifies the churn status. The ideal number of neurons within the hidden layer is established through an empirical, iterative process. Each hidden layer neuron embodies an intermediate prediction, receiving connections from a group of input layer neurons. These neuron outputs are subsequently processed by an activation function before being transmitted to the final output layer. In these architectures, the sigmoid function is commonly utilized as the primary activation function.

The MLP neural network is trained using the backpropagation algorithm, which fundamentally involves two distinct phases: forward propagation and backward error propagation. During forward propagation, input data traverses through the network layers towards the output, with each layer exclusively influencing the subsequent one. Should the network's predicted output deviate from the expected outcome, the error is computed and subsequently propagated backward through the network. This backward propagation facilitates the iterative adjustment of weights and thresholds until the network converges towards an optimal solution (Nielsen, 2015; Goodfellow et al., 2016; Géron, 2022). Upon definition of the network parameters, a validation methodology is implemented to rigorously test and evaluate the designed neural network's performance on the problem data. Among various validation techniques, the 70-30% split method is a common and extensively utilized approach, where 70% of the dataset is allocated for training and the remaining 30% for testing. Crucially, the dataset was split once into training and testing subsets using a fixed random seed (random_state), and this same split was used consistently across all model runs (for the proposed method, the simple MLP, and the SVM) to ensure a fair and reproducible performance comparison. Therefore, the intrinsic nature of neural network-based problem-solving inherently relies on this bipartite division of the dataset into training and testing subsets (Géron, 2022).

The training process and computations within the neural network's neurons and layers are governed by a fundamental input-output equation. This equation, formally presented as Equation (7), defines the aggregation of weighted inputs and thresholds within a neuron:

$$y_p^k = logit_p^k \left[\sum_{i=1}^{N_{k-1}} W_{ip}^{k-1} . y_i^{k-1} - \beta_i^k \right], p = 1,2, .. N_k; \;\; k = 1,2, .., M \tag{7}$$

Specifically, the connection weights between the i-th neuron in layer $k-1^{ip}$ and the p-th neuron in layer k is represented by W_{ip}^{k-1}. The output of the p-th neuron in layer k is denoted as y_p^k. The logistic activation function for the

p-th neuron in layer k is referred to as sgm_p^k, while β_i^k is the corresponding threshold. The logistic activation function, which is critical for introducing non-linearity into the network, is defined by Equation (8) as follows:

$$logit(a) = \frac{1}{1 + e^{-a}} \tag{8}$$

MLP Neural Networks have been extensively utilized for solving a diverse array of scientific and engineering problems, a widespread adoption facilitated by the advent of effective learning algorithms such as backpropagation. In the context of this study, the generalized delta rule is employed as the primary learning mechanism, whereby network weights are iteratively updated based on the discrepancies between the network's computed output and the desired target output. This iterative optimization process is commonly referred to as the backpropagation algorithm.

The training protocol for the backpropagation algorithm adheres to the following structured steps:

1. Initialize all weights randomly.
2. Compute the output vector.
3. Calculate the propagation error.
4. Update weights using Equation (9).
5. Compute the total error (ε) using Equation (11).
6. Repeat the algorithm from step 2 onward until the convergence condition is met. (i.e., achieving an error close to zero).

$$W_{ip}^{k-1}(t+1) = W_{ip}^{k-1}(t) + \alpha \sum_{n=1}^{l} \delta_{np}^{k} y_{ni}^{k-1} \tag{9}$$

Where *t*, represents the iteration of the algorithm, and α is the learning rate.

$$\delta_{np}^{k} = logit_{np}^{k} \left[\sum_{t=1}^{N_{k+1}} \delta_{nl}^{k+1} W_{pt}^{k}\,(t) \right] \tag{10}$$

$$\varepsilon = \sum_{n=1}^{l} \sum_{j=1}^{N_M} \left(y_{nj}^{M} - \hat{y}_{nj}^{M} \right)^2 \tag{11}$$

IV. RESEARCH FINDINGS

This section delineates the empirical implementation and subsequently presents the comprehensive results derived from the proposed methodology. A comparative analysis, contrasting the performance of the developed model with conventional approaches, is thoroughly discussed at the culmination of this section. The computational implementation of the proposed algorithm was executed within the MATLAB software environment, employing the Teldata dataset for rigorous validation and verification of the method's efficacy.

A. Dataset

This research seeks to provide definitive evidence for the enhanced performance of a neural network model optimized with the AOS algorithm. The model's superiority is specifically benchmarked against traditional machine learning methods, such as conventional MLPs and SVMs, for the task of customer churn prediction using the Teldata dataset.

The Teldata dataset, a publicly available resource obtained from Kaggle.com, encompasses a substantial volume of data, specifically 7043 customer instances, each characterized by 20 distinct variables or features.

Before model training, the dataset underwent essential preprocessing, including normalization and rescaling to binary values (0,1). The target variable, 'Churn', is a binary classification output (0 for 'No Churn' and 1 for 'Churn'). An analysis of the dataset revealed a notable class imbalance: approximately 73.46% of the 7042 instances belong to the 'No Churn' class, while 26.54% represent the 'Churn' class. This imbalance is common in real-world customer churn datasets. It is important to note that robustness of the model's performance was evaluated using classification metrics suitable for imbalanced datasets, as detailed in Section IV.B. Table II provides a detailed enumeration of these 20 features, specifying 19 independent variables and one dependent variable pertinent to the problem.

For numerical variables (e.g., tenure, MonthlyCharges, TotalCharges), the Mean represents the average value. For binary variables (e.g., SeniorCitizen, Partner, Dependents, PhoneService, PaperlessBilling, Churn), the Mean represents the proportion of instances coded as '1', and the Mode (Frequency) indicates the count/percentage of the most frequent category. For other categorical variables (e.g., gender, MultipleLines, InternetService), the Mode (Frequency) indicates the most frequent category and its percentage, while Mean and Standard Deviation are not provided because they are not conceptually meaningful for nominal data. Min and Max values represent the numerical encoding range.

Table II. Summary of Problem Variables

Variable Name	Variable Type	Min	Max	Mean / Mode (Frequency)	Std Dev
gender	Independent (Binary)	0	1	0 (50.3%)	N/A
SeniorCitizen	Independent (Binary)	0	1	0.1621 (16.2%)	0.3686
Partner	Independent (Binary)	0	1	0.4830 (48.3%)	0.4997
Dependents	Independent (Binary)	0	1	0.2995 (29.9%)	0.4581
tenure	Independent (Numerical)	0	72	32.3787	24.5574
PhoneService	Independent (Binary)	0	1	0.9032 (90.3%)	0.2957
MultipleLines	Independent (Categorical)	0	2	0 (48.4%)	N/A
InternetService	Independent (Categorical)	0	2	1 (43.9%)	N/A
OnlineSecurity	Independent (Categorical)	0	2	0 (49.6%)	N/A
OnlineBackup	Independent (Categorical)	0	2	0 (43.9%)	N/A
DeviceProtection	Independent (Categorical)	0	2	0 (43.9%)	N/A
TechSupport	Independent (Categorical)	0	2	0 (49.4%)	N/A
StreamingTV	Independent (Categorical)	0	2	0 (39.0%)	N/A
StreamingMovies	Independent (Categorical)	0	2	0 (38.8%)	N/A
Contract	Independent (Categorical)	0	2	0 (55.0%)	N/A
PaperlessBilling	Independent (Binary)	0	1	0.5921 (59.2%)	0.4915
PaymentMethod	Independent (Categorical)	0	3	0 (33.5%)	N/A
MonthlyCharges	Independent (Numerical)	18.25	118.75	64.7584	30.091
TotalCharges	Independent (Numerical)	18.8	8684.8	2280.91	2268.03
Churn	Dependent (Binary)	0	1	0.2654 (26.5%)	0.4416

B. Numerical results

This investigation rigorously analyzes the intricate relationships between the independent variables (the 19 features delineated in Table II) and their profound influence on the prediction of the dependent variable, namely the customer churn rate. Beyond mere prediction, the proposed optimized MLP methodology elucidates the precise contribution and influence of each independent variable on customer churn. This analytical capability is pivotal for identifying the most significant data features impacting churn, thereby serving as an invaluable decision-support tool for telecommunication company managers. By strategically prioritizing and addressing these critical factors, companies can significantly enhance service quality and fortify customer retention initiatives. Table III presents the adjustment parameters utilized within the proposed diagnostic method.

Table III. Summary of Adjustment Parameters Utilized in the Proposed Diagnostic Method

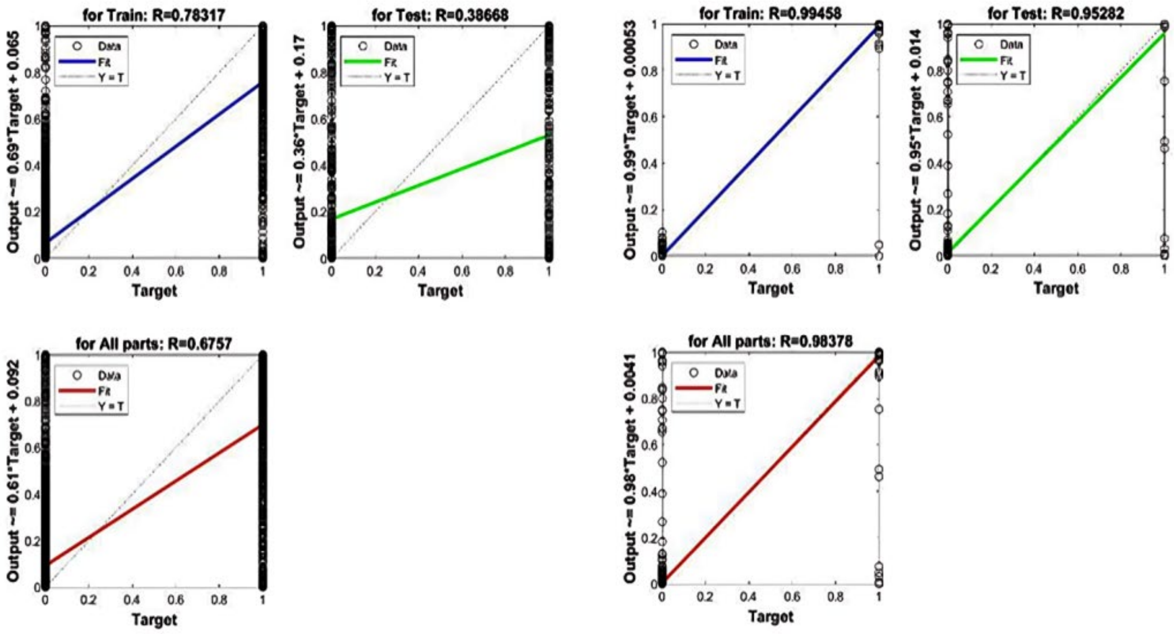
Neural Classifier	Number of Epochs	1000
	Cost Type	MSE (Mean Squared Error)
	Hidden Layer Neuron Range	30-1
	Training Algorithm Type	Bayesian Backpropagation
	Data Split for Training and Testing	70% for Training 30% for Testing
Atomic Orbital Optimization	Number of Solutions	20
	Number of Objectives	1
	Maximum Iterations	100
	Parameter R	0.5
	Alpha and Beta	Random
	Phi and Gamma	Random

The configuration of the neural network's architecture, specifically the number of neurons in the input, hidden, and output layers, is critically determined for optimal problem prediction. While the input layer neuron count directly corresponds to the number of data features in the problem, and the output layer neurons align with the problem's class label values, the optimal number of hidden layer neurons is typically ascertained through an iterative trial-and-error process. In this study, hidden layer neurons were systematically tested within a range of 1 to 20 to empirically identify the most effective configuration. Empirical findings revealed that employing 20 neurons within the hidden layer, coupled with Bayesian backpropagation training, yielded a remarkably high classification accuracy for the proposed methodology. For the training phase, the *trainbr* command, integrated within the neural network toolbox of MATLAB, was utilized for the Bayesian backpropagation algorithm.

It is pertinent to note that this particular training method inherently partitions the dataset into only two subsets, a training set and a test set, thereby precluding a distinct validation phase. Consequently, performance metrics, including classification metrics and confusion matrices, were exclusively evaluated for these training and test samples.

Classification metrics were employed to rigorously assess the robustness of the proposed method by examining the interrelationships among features to predict the binary output states of customer churn or non-churn. As depicted in Fig. 3, parts (a) and (b) provide a direct visual comparison between the simple (non-optimized) neural network method and the proposed optimized MLP approach, respectively. The graphical representations unequivocally illustrate the superior accuracy of the proposed optimized MLP method compared to its non-optimized counterpart.

Furthermore, the optimized MLP method, configured with 21 neurons in its hidden layer, consistently demonstrated the most favorable response and achieved optimal classification, leading to highly accurate output predictions with minimal error.



a. Simple MLP Method (Without Optimization) **b. Proposed Optimized MLP Method**

Fig. 3. Performance Comparison of Optimized and Non-Optimized MLP Models

As previously mentioned, Fig. 3 effectively elucidates the analytical prowess of the methodologies in discerning relationships among the independent variables under investigation. Specifically, Fig. 3(a) illustrates accuracies of approximately 99% and 98% for the training, testing, and combined datasets, respectively, for the proposed optimized method. In stark contrast, the conventional simple MLP method yielded significantly lower accuracies of approximately 78%, 38%, and 67% for the training, testing, and combined datasets, respectively.

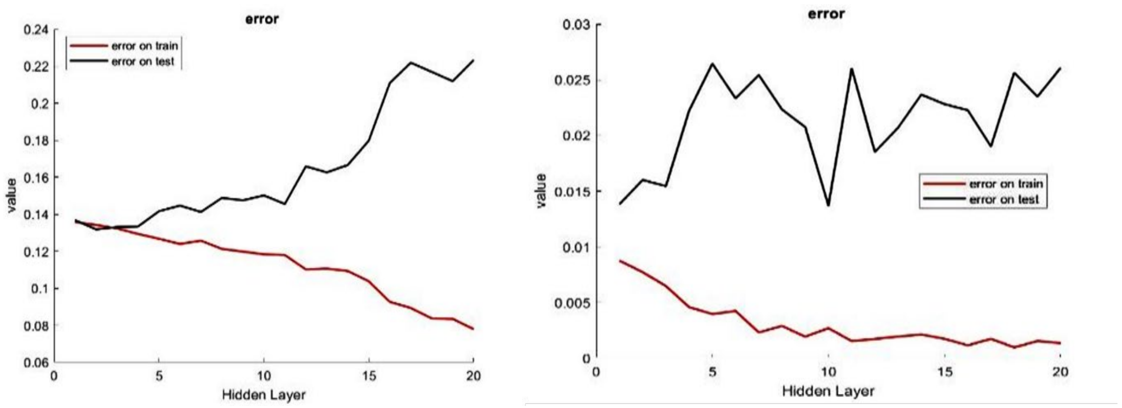
Further comprehensive analysis of predictive performance is presented in Fig. 4, which scrutinizes the outcomes of the proposed optimized MLP method versus the simple MLP method through the lens of confusion matrix analysis. The results delineate the optimal performance in terms of confusion, with the proposed optimized MLP method utilizing 21 hidden layer neurons and the simple MLP method employing 20 hidden layer neurons. The insights provided by the rightmost column of the confusion matrix explicitly detail the percentages of accurately and inaccurately classified samples for each output class, commonly referred to as the correct and incorrect prediction rates. As visibly demonstrated in Fig. 4, the test sample section exhibits notably high percentages of correctly classified predicted samples across all classes.

Fig. 4 provides a direct comparative analysis of the predictive efficacy between the proposed optimized MLP method and its simple (non-optimized) counterpart, utilizing the confusion matrix as the evaluation criterion. As previously elaborated, confusion analysis is instrumental in assessing a model's proficiency in accurately identifying both correctly and incorrectly predicted data instances.

a. Simple MLP Method (Without Optimization) **b. Proposed Optimized MLP Method**

Fig. 4. Confusion Estimation from Optimized and Non-Optimized MLP Prediction Methods

An algorithm's performance can be visualized in a confusion matrix, which is a tabular representation where each row corresponds to the actual class of an instance, and each column shows the class predicted by the model. The diagonal entries of the matrix indicate the number of accurate predictions, while the off-diagonal entries represent misclassifications. This matrix serves as the basis for deriving several key metrics: True Positives (TP) are positive instances correctly identified, True Negatives (TN) are negative instances correctly identified, False Positives (FP) are negative instances incorrectly labeled as positive, and False Negatives (FN) are positive instances misclassified as negative.



a Simple MLP Method (Without Optimization) **b. Proposed Optimized MLP Method**

Fig. 5. Error Estimation for Training and Testing Phases in Optimized and Non-Optimized MLP Prediction Methods

Specifically, Fig. 4(a) illustrates that the proposed optimized MLP method achieved remarkable confusion accuracies of 99.8%, 97.8%, and 99.3% for the training, testing, and combined datasets, respectively. In stark contrast, the simple MLP method exhibited lower confusion accuracies, registering 75.7%, 91.9%, and 87.7% for the respective training, testing, and combined datasets.

These compelling results unequivocally demonstrate that the proposed method, significantly enhanced through optimization, exhibits superior capability for analyzing complex relationships among independent variables to predict churn rates with greater precision, particularly compared to a traditional neural network with randomly initialized weights. This outcome substantially bolsters confidence in the robustness and efficacy of the developed approach. Further insights into performance are provided by Fig. 5, which delineates the error and accuracy metrics for both the training and testing phases. These results are derived from the application of both the proposed optimized method and the simple MLP on the dataset, specifically for training executed via the Bayesian backpropagation algorithm across a hidden layer range of 1 to 20 neurons.

Fig. 5 provides a visual representation of the training and testing errors for both the proposed optimized method (depicted in part b) and the simple, non-optimized MLP method (shown in part a). A meticulous examination of Fig. 5 reveals a discernible difference in error ranges. For the proposed method, the training and testing errors consistently reside within a narrow interval of [0-0.03]. Conversely, the comparatively simple MLP method exhibits a substantially broader error range, spanning [0.06-0.24]. This disparity is particularly illustrative, indicating that even the lower bound of the simple method's error range exceeds the upper limit observed in the proposed methodology. The significantly reduced error range achieved by the proposed optimized MLP method, relative to its non-optimized counterpart, robustly enhances the reliability and trustworthiness of the obtained churn prediction findings.

The calculation of the churn prediction error systematically involves a comparison of the model's estimated outputs with the actual data outputs across the training, testing, and combined datasets. For instance, to quantify the prediction error for the testing dataset, the outputs predicted by the proposed optimized MLP for each data point within the test set are rigorously juxtaposed against their corresponding true labels. The aggregate number of misclassified outputs is then divided by the total count of test data points to yield the error rate. Subsequently, the classifier's accuracy is determined by converting this error into a percentage and subtracting it from 100%. Furthermore, it is noteworthy that the analytical framework could be extended to explore repeated classification tasks through other advanced machine learning paradigms, such as ensemble learning, which inherently combines the outputs of multiple algorithms where the prediction of one serves as input for the next.

To provide a more comprehensive evaluation of classification performance, particularly given the observed class imbalance, a full suite of metrics, including Accuracy, Precision, Recall, and F1-Score, was calculated for both Class 0 ('No Churn') and Class 1 ('Churn') on the test set. These results, along with the overall accuracy, are summarized in Table IV.

As evident from Table IV, the Proposed Method consistently outperforms the Simple MLP across all evaluated metrics. The SVM's performance metrics indicate competitive overall accuracy but demonstrate varied class-wise performance. This highlights the effectiveness of the proposed optimized MLP-AOS model in achieving high, balanced predictive performance for customer churn.

In response to the need for greater model interpretability, a feature importance analysis was performed to identify the most influential factors affecting customer churn. Given the complex architecture of the optimized MLP, a permutation importance analysis was conducted on the model's outputs to robustly identify the key features driving the predictions. The results of this analysis, which rank the features based on their contribution to the model's predictive power, are summarized in Table V. As shown in the table, the most critical features are primarily related to a customer's contractual and financial relationship with the company. The analysis revealed that tenure, TotalCharges, and MonthlyCharges are the top three most influential features. These findings offer practical insights for telecommunication managers, enabling them to formulate targeted retention strategies by focusing on customers with short tenure, low total charges, or specific contractual arrangements.

Table IV. Classification Performance Metrics on Test Set

Model	Criteria	0 (No Churn) Class	Class 1 (Churn)	Overall
Optimized MLP-AOS	Accuracy	-	-	0.9782
	Precision	0.9889	0.9483	-
	Recall	0.9817	0.9684	-
	F1-Score	0.9853	0.9582	-
Simple MLP	Accuracy	-	-	0.757
	Precision	0.8311	0.6236	-
	Recall	0.8911	0.4991	-
	F1-Score	0.8601	0.5545	-
Support Vector Machine	Accuracy	-	-	0.7932
	Precision	0.8301	0.6462	-
	Recall	0.9034	0.4884	-
	F1-Score	0.8652	0.5563	-

Table V. Feature Importance Scores

Feature Name	Importance Score
Tenure	0.2188
Total Charges	0.1837
Monthly Charges	0.1264
Contract	0.1039
Internet Service	0.0766
Payment Method	0.0622
Online Security	0.0353
Tech Support	0.0344
Online Backup	0.0249
MultipleLines	0.0238
Partner	0.0175
Device Protection	0.0172
Gender	0.0156
Dependents	0.0142
Paperless Billing	0.0139
Streaming TV	0.0125
Streaming Movies	0.0116
Senior Citizen	0.0094
Phone Service	0.0039

V. DISCUSSION AND CONCLUSIONS

The present study addressed the critical challenge of customer churn prediction in the telecommunications sector by introducing a new hybrid methodology. This approach strategically integrates the Multilayer Perceptron (MLP) neural network with the Atomic Orbital Search (AOS) metaheuristic algorithm to overcome the inherent limitations of traditional MLPs. The empirical findings, derived from the Teldata dataset, provide strong evidence for the efficacy and robustness of the proposed model.

A key finding is the superior predictive performance achieved by the optimized MLP-AOS model. With an impressive 97.8% accuracy on the test set, the model not only outperformed a conventional simple MLP but also demonstrated a highly competitive performance compared to other recent comparable studies. For instance, while Sagming et al. (2025) reported an accuracy of 98.50% using a more complex TDA-based XGBoost model, our model's performance is highly comparable, especially given its robust performance on the imbalanced dataset. This result underscores the effectiveness of leveraging a specialized metaheuristic, such as AOS, to fine-tune MLP parameters, a challenge noted in prior work (Adwan et al., 2014).

Analysis of key classification metrics, including Precision, Recall, and F1-Score, further demonstrates the model's strength in handling class imbalance. By achieving strong performance on both the majority ('No Churn') and minority ('Churn') classes, the model mitigates the risk of a biased classification—a common limitation highlighted in the literature (Chai et al., 2025). The feature importance analysis revealed that contractual and financial variables such as tenure, TotalCharges, and MonthlyCharges are the most critical determinants of churn. This finding is consistent with general observations in the field and provides actionable insights for managers to focus their retention efforts on these key customer segments (Boobier, 2018).

Our work is further contextualized within the latest advancements in the field. The recent comprehensive review by Manzoor et al. (2024) highlighted the need for more interpretable and profit-based evaluation metrics in churn prediction, which our model partially addresses through feature importance analysis. Studies by Maskale et al. (2024) on bank customer churn and Bhushan et al. (2024) on developing explainable models for telecom churn reflect a growing industry focus on practical, domain-specific, and interpretable solutions. The performance and framework of our proposed model align with these emerging trends, showcasing its relevance and timeliness.

Indeed, a broader trend across both industrial engineering and business intelligence is the successful integration of new predictive and hybrid models to enhance estimation accuracy and support complex decision-making. This is further substantiated by other recent works, such as the study by Movafaghpour (2023) proposing a new regression tree method for addressing project delays, the research by Abbaspour Ghadim Bonab et al. (2022) using a predictive data-driven approach with a machine learning algorithm (ANFIS) to solve a decision problem for an inventory system, and the work by Behdinian et al. (2022) integrating a machine learning algorithm with a simulation method to improve software project management. In this vein, our hybrid MLP-AOS model for customer churn prediction aligns directly with this effective approach.

In addition to these findings, this study offers several significant contributions. First, it addresses a key research gap by providing an optimized framework that effectively mitigates the issues of arbitrary weight initialization and local optima in MLPs. Second, by demonstrating robust performance on a real-world, imbalanced dataset, our work highlights the practical applicability of advanced optimization techniques in a business context. Furthermore, the methodology, by re-evaluating an SVM model for comprehensive metrics, also sets a standard for thorough and fair model comparisons.

Despite these achievements, some avenues for future work remain. While the model's performance is strong, the computational complexity and execution time of advanced optimization methods like AOS can be a challenge (Chugh et al., 2022). Future research could explore more computationally efficient optimization strategies or hybrid models to reduce training time. Furthermore, while we provided a feature importance analysis, integrating other interpretability methods or extending the analysis to a wider range of datasets would be a valuable direction for future studies. The adoption of other intelligent methodologies, such as ensemble learning, or the application of Adaptive Neuro-Fuzzy Inference Systems (ANFIS), also presents promising avenues to explore for further enhancing churn prediction capabilities.

REFERENCES

Abbaspour Ghadim Bonab, A., & Yousefi Nejad Attari, M. (2022). A predictive data-driven state-dependent decision approach to determine inventory system states for critical spare parts. *Journal of Quality Engineering and Production Optimization*, *1*(2), pp.1–13.

Adwan, O., Faris, H., Jaradat, K., Harfoushi, O., & Ghatasheh, N. (2014). Predicting customer churn in the telecom industry using multilayer perceptron neural networks: Modeling and analysis. *Journal of Computer Science and Technology*, *11*(3), pp.75–81.

Alsiehemy, A. (2019). An Assessment of Customer Retention with Self-Service Technologies: A Model Development. *Business Management and Strategy*, *10*(1), pp.78–92.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunications sector using a rough set approach. *Neurocomputing*, *237*, pp.242–254.

Azizi, M. (2020). Atomic orbital search: A novel metaheuristic algorithm. *Applied Mathematical Modelling*, *89*(1), pp.123–134.

Azizi, M., Talatahari, S., Khodadadi, N., & Sareh, P. (2022). Multiobjective Atomic Orbital Search (MOAOS) for Global and Engineering Design Optimization. *IEEE Access*, *10*, pp. 67727–67746.

Behdinian, A., Amani, M. A., Aghsami, A., & Jolai, F. (2022). An Integrating Machine Learning Algorithm and Simulation Method for Improving Software Project Management: A Case Study. *Journal of Quality Engineering and Production Optimization*, *1*(2), pp.1–11.

Bhushan, S. B. (2024). *Enhancing Customer Churn Prediction in the Telecom Sector Using Advanced Machine Learning Techniques and Explainable AI* [Master's Research Project, National College of Ireland]. National College of Ireland Institutional Repository (NORMA). https://norma.ncirl.ie/8683/

Bi, W., Cai, M., Liu, M., & Li, G. (2016). A big data clustering algorithm for mitigating the risk of customer churn. *IEEE Transactions on Industrial Informatics*, *12*(3), pp.1270–1281.

Boobier, T. (2018). Advanced analytics and AI: Impact, implementation, and the future of work. Hoboken, NJ: John Wiley & Sons.

Chai, E., Khadullo, K., & Tole, K. (2025). Enhancing customer churn prediction: Addressing disparities and imbalances in machine learning models. *Journal of Machine Learning Research*, *13*(3), pp.129–148.

Chugh, A., Sharma, V. K., Bhatia, M. K., & Jain, C. (2022). A big data query optimization framework for telecom customer churn analysis. In A. Khanna, S. Bhattacharyya, S. Anand, D. Gupta, A. E. Hassanien, & A. Jaiswal (Eds.), *Advances in Intelligent Systems and Computing* (pp. 475–484). Singapore: Springer.

Deng, W., Deng, L., Liu, J., & Qi, J. (2019). Sampling method based on an improved C4.5 decision tree and its application in the prediction of telecom customer churn. *International Journal of Information Technology and Management*, *18*(1), pp.93–109.

Ever, Y. K., Dimililer, K., & Sekeroglu, B. (2019). Comparison of machine learning techniques for prediction problems. In L. Barolli, M. Takizawa, F. Xhafa, & T. Enokido (Eds.), *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019)* (pp. 713–723). Matsue, Japan: Springer.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.)*. Sebastopol, CA: O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

Han, H., Back, K.-J., & Barrett, B. (2009). Influencing factors on restaurant customers' revisit intention: The roles of emotions and switching barriers. *International Journal of Hospitality Management*, *28*(4), pp.563–572.

Huettmann, F., Craig, E. H., Herrick, K. A., Baltensperger, A. P., Humphries, G. R. et al. (2018). Use of machine learning (ML) for predicting and analyzing ecological and 'presence only' data: An overview of applications and a good outlook. In F. Huettmann, D. J. Lieske, S. N. R. Mohseni, E. H. Craig, & S. C. L. Li (Eds.), *Machine Learning for Ecology and Sustainable Natural Resource Management* (pp. 27–61). Cham: Springer.

Hossein Khani, N., Hosseini Motlagh, S. M., & Khakzar Bafruei, M. (2014). *Identification of factors affecting customer churn in the insurance industry.* Paper presented at the 21st National and 7th International Insurance and Development Conference, Tehran, Iran. SID. https://sid.ir/paper/833543/fa

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2021). Customer churn prediction system: a machine learning approach. *Computing*, *104*, pp.271–294.

Lee, J., Lee, J., & Feick, L. (2001). The impact of switching costs on the customer satisfaction-loyalty link: Mobile phone service in France. *Journal of Service Management*, *15*(1), pp.35–48.

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners. *IEEE Access*, *12*, pp.70434–70445.

Maskale, V., Vaidya, V., Patil, Y., Bagal, Y., & Dhamdhre, V. (2024, April). To Design and implement an Application for Bank Customer Churning Rate Prediction and Analysis using a Machine Learning Algorithm. In *2024, MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)* (pp. 475–484). Pune, India: IEEE.

Movafaghpour, M. A. (2023). Predicting Project Delays Using a New Trended Regression Tree Method. *Journal of Quality Engineering and Production Optimization*, *2*(1), pp.29–41.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, *43*(2), pp.204–211.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.

Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2017). *Human resource management: Gaining a competitive advantage*. New York, NY: McGraw-Hill Education.

Pustokhina, I. V., Pustokhin, D. A., Aswathy, R., Jayasankar, T., Jeyalakshmi, C., & Díaz, V. G. (2021). Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. *Computers & Industrial Engineering*, 162, 107606.

Rai, S., Khandelwal, N., & Boghey, R. (2020). Analysis of customer churn prediction in the telecom sector using the CART algorithm. In A. K. Luhach, J. A. Kosa, R. C. Poonia, X.-Z. Gao, & D. Singh (Eds.), *Sustainable Technologies for Computational Intelligence: Proceedings of ICSTCI 2020* (AISC, Vol. 1045, pp. 457–466). Ghaziabad, India: Springer.

Ramesh, P., Jeba Emilyn, J. & Vijayakumar, V. (2022). Hybrid Artificial Neural Networks Using Customer Churn Prediction. *Wireless Personal Communications*, *124*, pp.1695–1709.

Rani, K. S., Thaslima, S., Prasanna, N. G., Vindhya, R., & Srilakshmi, P. (2021). Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression. *International Journal of Innovative Research in Computer Science & Technology*, *9*(4), pp.2347–5552.

Ren, C. R., Hu, Y., & Cui, T. (2019). Responses to rival exit: Product variety, market expansion, and preexisting market structure. *Strategic Management Journal*, *40*(2), pp.253–276.

Sagming, M., Heymann, R., & Visaya, M. V. (2025). Using topological data analysis and machine learning to predict customer churn. *Journal of Big Data*, *12*(4), pp.320–338.

Sato, T., Huang, B. Q., Huang, Y., Kechadi, M.-T., & Buckley, B. (2010). Using PCA to predict customer churn in a telecommunication dataset. In L. Cao, Y. Liu, C. Zhang, & H. Zhu (Eds.), *Advanced Data Mining and Applications: 6th International Conference, ADMA 2010* (Lecture Notes in Computer Science, Vol. 6460, pp. 198–208). Chongqing, China: Springer.

Zhang, T., Moro, S., & Ramos, R. F. (2025). A data-driven approach to improve customer churn prediction based on telecom customer segmentation. *Journal of Machine Learning Research*, *13*(5), pp.213–229.