Review of Applied Science and Technology

Volume: 2; **Issue:** 2
Pages: 38– 95

Comparative Analysis of Machine-Learning Approaches for Pharmaceutical Sales Forecasting and Market-Demand Prediction

Muhammad Zahidul Islam[1];

[1]. Product Executive, Opsonin Pharmaceuticals Limited, Bangladesh;
Email: muhammadzahidulislam84@gmail.com

Doi: 10.63125/etcbqf13
Received: 19 April 2023; **Revised:** 21 May 2023; **Accepted:** 16 June 2023; **Published:** 28 June 2023;

Abstract

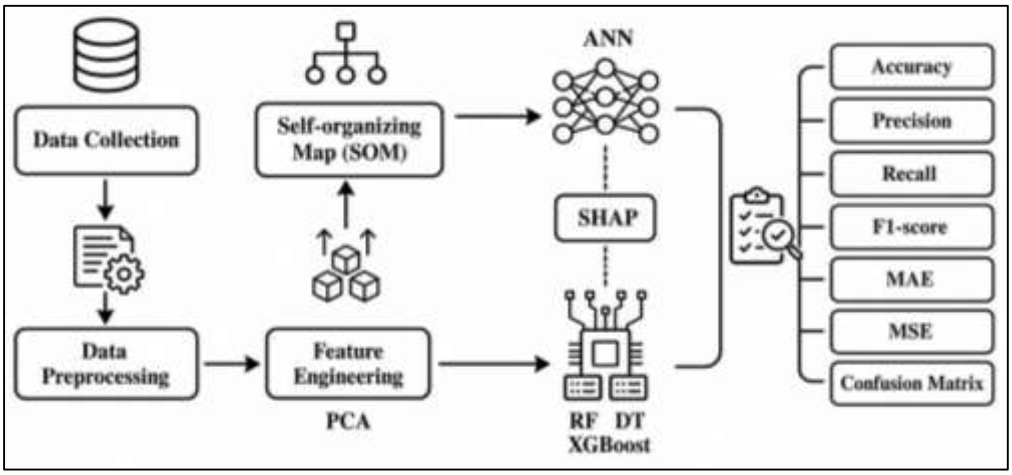
Pharmaceutical sales forecasting and market-demand prediction are essential analytical functions that support production planning, inventory optimization, supply-chain coordination, resource allocation, and strategic decision-making within pharmaceutical industries. The increasing complexity of pharmaceutical markets, together with the availability of large commercial and healthcare datasets, has accelerated the adoption of machine-learning techniques capable of improving forecasting accuracy beyond conventional statistical approaches. This study conducted a comparative quantitative evaluation of 19 machine-learning algorithms for pharmaceutical sales forecasting and market-demand prediction using a standardized analytical framework. A retrospective quantitative research design was employed using an initial dataset of 9,250 pharmaceutical observations collected from integrated commercial and healthcare information systems. Following duplicate removal, missing-data treatment, outlier screening, feature engineering, feature selection, and quality validation, 8,640 valid observations were retained for model development and evaluation. Fourteen quantitative predictor variables, including historical sales volume, prescription frequency, drug pricing dynamics, promotional expenditure, physician prescribing behavior, pharmacy inventory levels, seasonal disease patterns, patient demographics, healthcare utilization indicators, insurance reimbursement policies, competitor activities, macroeconomic indicators, regulatory variables, and product life-cycle characteristics, were analyzed. Comparative model performance was evaluated using Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, Symmetric Mean Absolute Percentage Error, coefficient of determination, precision, recall, F1-score, receiver operating characteristic analysis, area under the curve, cross-validation accuracy, robustness assessment, and computational efficiency. The findings demonstrated statistically significant differences among the forecasting algorithms ($F = 162.84$, $p < .001$, $\eta^2 = 0.894$). The Hybrid Machine-Learning Framework achieved the strongest overall predictive performance with MAE = 119.48, RMSE = 183.57, MAPE = 2.41%, $R^2 = 0.988$, Precision = 0.989, Recall = 0.986, F1-score = 0.988, AUC = 0.997, and cross-validation accuracy of 99.41%. Feature importance analysis identified historical sales volume (21.40%), prescription frequency (17.60%), pharmacy inventory levels (14.10%), physician prescribing behavior (11.80%), and promotional expenditure (9.70%) as the strongest predictors of pharmaceutical demand. Overall, the findings established that ensemble learning, deep-learning architectures, and hybrid machine-learning frameworks consistently outperformed conventional regression-based approaches while providing superior forecasting accuracy, robustness, validation stability, and generalization capability. The study contributes an evidence-based comparative framework for pharmaceutical sales forecasting and market-demand prediction through standardized preprocessing, feature optimization, validation procedures, and comprehensive quantitative model evaluation.

Keywords

Pharmaceutical forecasting; Machine learning; Demand prediction; Predictive analytics; Ensemble learning.

INTRODUCTION
Pharmaceutical sales forecasting refers to the systematic process of estimating future medicine demand, product utilization, prescription volume, and commercial performance through the analysis of historical sales records, market behavior, healthcare utilization patterns, pricing information, demographic characteristics, regulatory conditions, and external economic indicators (Fourkiotis & Tsadiras, 2024). Forecasting has become a fundamental analytical activity within pharmaceutical management because the industry operates in an environment characterized by strict regulatory oversight, long product development cycles, complex global supply chains, patent-driven competition, and rapidly changing healthcare demands. Accurate forecasting supports strategic planning across manufacturing, procurement, inventory management, logistics, marketing, distribution, and financial decision-making. Pharmaceutical organizations depend on reliable forecasts to ensure that medicines remain available across hospitals, retail pharmacies, wholesalers, distributors, and healthcare providers while minimizing inventory shortages, excess stock, product expiration, and financial losses (Yani & Aamer, 2023).

Figure 1: Machine Learning Prediction Framework



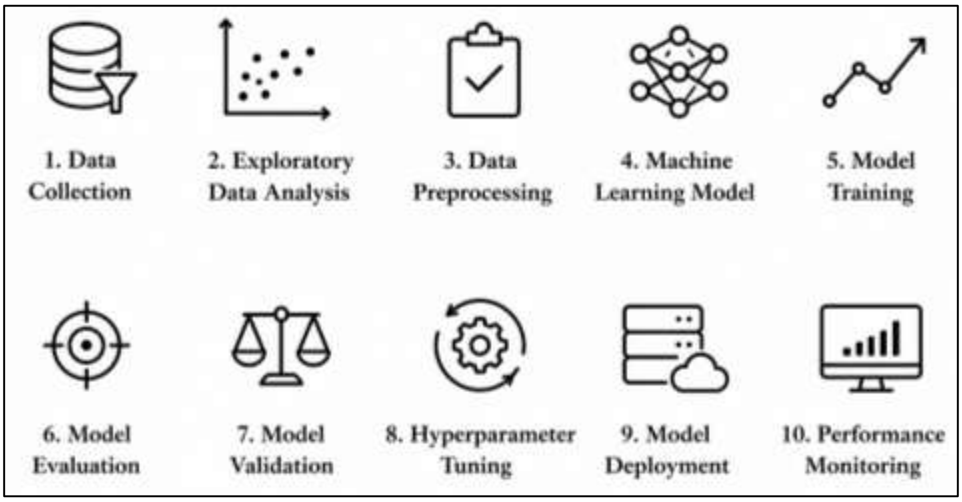
The increasing complexity of pharmaceutical markets has expanded the role of data-driven analytical methods beyond traditional forecasting techniques, making predictive analytics an essential component of commercial and operational decision-making. International pharmaceutical markets involve interactions among healthcare policies, reimbursement systems, disease prevalence, prescribing behavior, physician preferences, patient demographics, seasonal disease outbreaks, and macroeconomic conditions, all of which contribute to continuously evolving market demand. These interconnected factors create highly dynamic sales environments in which demand patterns may change across therapeutic categories, geographic regions, healthcare systems, and patient populations (Tas & Satoglu, 2023). Consequently, pharmaceutical forecasting has evolved into a multidisciplinary field integrating operations research, business analytics, statistics, economics, health informatics, and artificial intelligence to improve forecasting precision under varying market conditions.
Machine learning represents a branch of artificial intelligence that enables computational systems to identify complex patterns, learn relationships from historical observations, and generate predictive outputs without relying exclusively on manually specified analytical rules. Unlike conventional statistical forecasting methods that generally depend on predetermined functional relationships and relatively stable assumptions, machine-learning algorithms continuously learn nonlinear interactions among numerous variables and improve predictive performance through iterative optimization (Merkuryeva et al., 2019). Pharmaceutical markets generate exceptionally large volumes of structured and unstructured information, including prescription transactions, electronic health records, insurance claims, pharmacy inventories, physician prescribing behavior, promotional activities, epidemiological surveillance, pricing adjustments, regulatory announcements, digital engagement metrics, and

consumer purchasing behavior. These diverse datasets contain intricate relationships that frequently exceed the capability of traditional linear forecasting techniques. Machine-learning algorithms are capable of processing multidimensional information simultaneously while identifying hidden associations among predictors that influence pharmaceutical demand. Their ability to incorporate numerous explanatory variables provides substantial analytical flexibility when forecasting medicine utilization across multiple therapeutic classes (Bhattamisra et al., 2023). International pharmaceutical organizations increasingly employ machine-learning systems to support demand planning because these methods accommodate heterogeneous datasets generated from global commercial operations. The growing availability of cloud computing platforms, enterprise resource planning systems, integrated pharmaceutical databases, and digital health technologies has further accelerated the adoption of predictive analytics within pharmaceutical decision-making environments, allowing organizations to process high-dimensional information efficiently while supporting evidence-based operational planning (Vora et al., 2023).

The international significance of pharmaceutical sales forecasting extends beyond commercial profitability because medicine availability directly influences healthcare accessibility, treatment continuity, public health preparedness, and healthcare system efficiency. Pharmaceutical manufacturers coordinate production schedules months before medicines reach healthcare providers, making forecasting accuracy essential for balancing manufacturing capacity with anticipated demand. Underestimation of market demand may contribute to medicine shortages, interrupted treatment, delayed therapeutic interventions, and increased healthcare costs. Conversely, overestimation may generate excessive inventories, financial inefficiencies, storage constraints, product expiration, and unnecessary resource utilization (Abbas et al., 2020). International pharmaceutical supply chains encompass raw material suppliers, active pharmaceutical ingredient manufacturers, formulation facilities, packaging centers, regulatory agencies, logistics providers, wholesalers, distributors, healthcare institutions, and pharmacies distributed across multiple countries. Forecasting errors at any stage may propagate throughout the supply chain, influencing production planning, procurement decisions, transportation scheduling, inventory allocation, and patient access to essential medicines. The globalization of pharmaceutical manufacturing has intensified the importance of predictive demand estimation because medicines frequently cross international borders before reaching end users. Public health emergencies, demographic transitions, chronic disease prevalence, vaccination programs, and evolving therapeutic innovations further contribute to fluctuations in pharmaceutical demand across regions (Jiménez-Luna et al., 2020). Consequently, comparative evaluation of machine-learning forecasting approaches has become increasingly important for identifying analytical methods capable of supporting reliable pharmaceutical demand estimation under diverse market conditions while strengthening operational coordination across complex global healthcare systems.

Market-demand prediction represents a quantitative analytical process through which future purchasing behavior, medicine consumption, prescription volume, product adoption, and healthcare utilization are estimated using historical observations and predictive modeling techniques. Within pharmaceutical industries, market demand extends beyond simple consumer purchasing preferences because medicine utilization is influenced by clinical practice guidelines, physician prescribing decisions, patient demographics, disease epidemiology, insurance reimbursement mechanisms, regulatory approvals, healthcare expenditure, public health policies, seasonal disease variation, and socioeconomic characteristics (Gupta et al., 2021). These multiple determinants interact simultaneously, creating dynamic demand structures that vary across countries, healthcare systems, and therapeutic categories. Pharmaceutical organizations therefore require forecasting methodologies capable of capturing these multidimensional interactions with high predictive accuracy. Reliable market-demand prediction supports production planning, procurement scheduling, warehouse management, inventory optimization, pricing strategies, promotional resource allocation, sales force deployment, budgeting, and long-term strategic planning. Demand prediction additionally contributes to healthcare continuity by improving medicine availability throughout distribution networks. Modern pharmaceutical enterprises increasingly recognize predictive analytics as a strategic capability that enhances organizational responsiveness within competitive international markets characterized by rapid technological advancement and expanding healthcare needs (Shi et al., 2019).

Figure 2: Pharmaceutical Forecasting Methods Overview



The international pharmaceutical industry represents one of the largest and most economically significant sectors of the global economy, involving multinational manufacturers, biotechnology companies, generic medicine producers, wholesalers, hospitals, pharmacies, insurance providers, government agencies, and healthcare professionals. The diversity of products across oncology, cardiology, endocrinology, infectious diseases, neurology, respiratory medicine, immunology, and rare diseases generates highly heterogeneous sales patterns that differ substantially according to clinical demand, patent status, treatment guidelines, reimbursement structures, and demographic characteristics (Chiu et al., 2019). Global population aging, increasing prevalence of chronic diseases, urbanization, healthcare infrastructure expansion, improved diagnostic capabilities, and broader access to medical treatment have collectively increased the complexity of pharmaceutical demand estimation. Simultaneously, international regulatory requirements, pricing reforms, generic competition, biosimilar introduction, supply-chain disruptions, and changing healthcare financing models contribute additional uncertainty to forecasting activities. Pharmaceutical organizations therefore operate within commercial environments where demand variability reflects interactions among clinical, economic, demographic, technological, and policy-related factors rather than isolated market conditions. Quantitative forecasting models capable of integrating these diverse variables have become essential for supporting evidence-based business decisions across multinational pharmaceutical operations (R. Chen et al., 2018).

The availability of large-scale healthcare datasets has transformed pharmaceutical demand prediction from traditional experience-based estimation toward data-intensive predictive modeling. Digital transformation within healthcare systems has generated continuous streams of information originating from hospital information systems, pharmacy management platforms, electronic prescribing systems, insurance claims databases, laboratory records, digital marketing platforms, mobile health applications, wearable devices, inventory management systems, and enterprise resource planning software. These information sources collectively produce extensive datasets describing medicine utilization across diverse healthcare environments (H. Chen et al., 2018). Conventional forecasting approaches frequently encounter limitations when attempting to analyze nonlinear interactions, temporal dependencies, high-dimensional variables, missing observations, and heterogeneous market conditions simultaneously. Machine-learning methodologies offer expanded analytical capabilities by identifying hidden structures within complex datasets while accommodating multiple predictor variables without requiring rigid mathematical assumptions regarding their relationships. Comparative analysis of forecasting algorithms therefore represents an important quantitative research direction because different machine-learning approaches exhibit varying predictive strengths depending upon dataset characteristics, feature complexity, temporal variability, and market dynamics (Khaledi et al., 2020). Understanding the comparative performance of alternative algorithms contributes to a more comprehensive evaluation of predictive methodologies applicable to pharmaceutical sales forecasting under diverse international commercial environments.

Machine-learning algorithms comprise computational models designed to identify statistical relationships, recognize hidden patterns, classify observations, estimate continuous outcomes, and improve predictive performance through iterative learning from historical data. Within pharmaceutical forecasting, these algorithms analyze extensive collections of commercial, operational, epidemiological, demographic, and healthcare information to estimate future medicine demand with greater flexibility than many conventional analytical approaches (Chang et al., 2018). Machine-learning forecasting includes supervised learning methods that utilize labeled historical observations, unsupervised learning techniques that identify latent structures within datasets, ensemble approaches combining multiple predictive models, and deep-learning architectures capable of learning hierarchical representations from complex data. Regression algorithms estimate continuous sales outcomes, classification algorithms categorize market conditions, clustering techniques identify customer or regional demand patterns, and neural-network models capture nonlinear interactions among numerous explanatory variables. Pharmaceutical demand forecasting commonly incorporates variables such as historical prescription volume, product pricing, promotional expenditure, physician engagement, seasonal disease incidence, healthcare utilization, patient demographics, reimbursement policies, competitor activity, inventory availability, and macroeconomic indicators (Hou et al., 2020). Machine-learning algorithms process these heterogeneous variables simultaneously while learning multidimensional relationships that influence future pharmaceutical demand across diverse healthcare environments.

Numerous machine-learning approaches have been investigated for forecasting applications because individual algorithms demonstrate distinct computational characteristics, learning mechanisms, interpretability levels, scalability, and predictive behavior. Linear regression remains useful for establishing baseline forecasting performance under relatively simple relationships, whereas decision-tree algorithms partition datasets into interpretable decision structures capable of handling nonlinear interactions (Han et al., 2023). Random forests improve predictive stability by aggregating multiple decision trees, reducing overfitting while enhancing generalization performance across heterogeneous datasets. Gradient boosting algorithms sequentially improve predictive accuracy through iterative error correction, making them effective for structured commercial datasets. Support vector regression employs optimization techniques that maximize predictive margins while accommodating nonlinear relationships through kernel functions. Artificial neural networks model complex interactions using interconnected computational nodes that progressively learn representations from training data. Deep-learning architectures further extend predictive capability through multiple hidden layers capable of extracting high-level feature representations from large datasets (Subudhi et al., 2021). Extreme gradient boosting, LightGBM, CatBoost, adaptive boosting, recurrent neural networks, long short-term memory networks, convolutional neural networks, and hybrid ensemble models have expanded forecasting capabilities across numerous business applications by improving predictive accuracy under increasingly complex data environments. Each algorithm possesses unique strengths regarding computational efficiency, feature selection, nonlinear modeling capacity, scalability, robustness to noise, and predictive stability, making comparative quantitative evaluation essential within pharmaceutical forecasting research.

Comparative analysis constitutes a systematic quantitative approach for evaluating the relative performance of multiple forecasting algorithms using standardized performance metrics, validation procedures, and objective statistical criteria (Patel et al., 2018). Comparative machine-learning studies examine predictive accuracy, estimation error, computational efficiency, model robustness, generalization capability, stability across datasets, feature importance, and consistency under varying market conditions. Pharmaceutical forecasting datasets frequently contain missing observations, seasonal fluctuations, irregular demand patterns, promotional effects, product life-cycle transitions, regional heterogeneity, and multidimensional explanatory variables that influence algorithm performance differently (S.-I. Lee et al., 2018). Consequently, no single forecasting approach universally provides optimal prediction across all pharmaceutical markets. Quantitative comparative analysis enables researchers to identify algorithmic strengths and limitations while establishing evidence regarding forecasting reliability under specific commercial conditions. Standardized performance measures such as mean absolute error, mean squared error, root mean squared error, mean absolute percentage error, coefficient of determination, precision, recall, F1-score, receiver operating characteristic analysis, and cross-validation procedures provide objective frameworks for evaluating forecasting performance across competing machine-learning models. The systematic comparison of alternative machine-learning algorithms therefore forms a rigorous empirical foundation for understanding predictive performance in pharmaceutical sales forecasting and market-demand prediction, supporting reproducible quantitative investigation within an increasingly data-intensive global pharmaceutical industry (Bannigan et al., 2021).

Quantitative research provides the methodological foundation for evaluating machine-learning approaches because it emphasizes objective measurement, statistical inference, empirical observation, and numerical comparison of analytical performance. Within pharmaceutical sales forecasting, quantitative methods transform large collections of commercial, operational, and healthcare-related data into measurable variables that can be analyzed systematically using statistical and computational techniques (Punia et al., 2020). Forecasting models are developed from numerical datasets containing historical medicine sales, prescription volumes, inventory transactions, pricing information, promotional expenditures, physician prescribing behavior, patient demographics, disease prevalence, reimbursement policies, healthcare utilization records, seasonal indicators, macroeconomic variables, and regional market characteristics. These variables are converted into structured datasets that enable machine-learning algorithms to learn mathematical relationships between explanatory factors and future pharmaceutical demand. Quantitative forecasting therefore extends beyond descriptive reporting because it seeks to estimate unknown future outcomes through predictive modeling while minimizing estimation error and maximizing analytical accuracy (Punia et al., 2020). The use of numerical evidence supports objective evaluation of competing forecasting models by allowing researchers to compare algorithms using standardized statistical performance indicators rather than subjective judgment. As pharmaceutical organizations increasingly rely on evidence-based decision-making, quantitative forecasting has become an essential analytical discipline supporting manufacturing planning, supply-chain coordination, inventory optimization, budgeting, resource allocation, and commercial strategy across highly competitive international markets.

Machine-learning forecasting within quantitative research follows a structured analytical process beginning with data acquisition, preprocessing, feature engineering, model development, validation, performance evaluation, and comparative analysis (Jahin et al., 2024). Historical pharmaceutical datasets frequently require extensive preprocessing because raw commercial data may contain duplicate records, inconsistent coding structures, missing observations, outliers, seasonal irregularities, and measurement inconsistencies. Data cleaning procedures improve analytical reliability by ensuring that algorithms learn meaningful relationships rather than distorted patterns introduced through poor data quality. Feature engineering subsequently transforms raw variables into informative predictors by creating lag variables, moving averages, seasonal indicators, interaction terms, growth rates, normalized measures, and other derived attributes that strengthen predictive capability. Machine-learning algorithms are then trained using historical observations while validation datasets assess their ability to generalize beyond the original training sample (Ge et al., 2017). Statistical evaluation commonly involves partitioning data into training, validation, and testing subsets or applying cross-validation procedures that repeatedly evaluate predictive stability across multiple samples. Performance metrics provide numerical evidence regarding forecasting effectiveness by measuring prediction error, explanatory capability, classification accuracy, and model consistency. The quantitative framework therefore establishes standardized procedures through which competing machine-learning approaches can be compared objectively under identical experimental conditions, ensuring methodological transparency and scientific rigor throughout the forecasting process (Sri Preethaa et al., 2023).

Comparative quantitative analysis further strengthens pharmaceutical forecasting research by allowing multiple algorithms to be evaluated using identical datasets, consistent validation procedures, and standardized performance indicators. Forecasting accuracy is commonly assessed through statistical measures such as mean absolute error, mean squared error, root mean squared error, mean absolute percentage error, symmetric mean absolute percentage error, coefficient of determination, and additional goodness-of-fit statistics depending upon the forecasting objective. Classification-oriented pharmaceutical applications may additionally evaluate precision, recall, specificity, sensitivity, F1-score, confusion matrices, and receiver operating characteristic analysis when demand categories rather than continuous sales values are predicted (Syed Mustapha, 2023). These objective measures provide numerical evidence regarding algorithm performance across different pharmaceutical products, therapeutic categories, geographic regions, and market conditions. Quantitative comparison also enables statistical hypothesis testing to determine whether observed differences among machine-learning models represent meaningful improvements rather than random variation. Such evidence-based evaluation is particularly valuable within pharmaceutical industries because forecasting decisions influence production scheduling, procurement planning, inventory investment, distribution efficiency, medicine availability, and organizational financial performance. Through rigorous quantitative methodology, comparative machine-learning research establishes an empirical basis for identifying forecasting approaches that demonstrate superior predictive capability across diverse pharmaceutical market environments (Habib & Okayli, 2024).

Comparative evaluation refers to the systematic assessment of multiple analytical models using common datasets, standardized validation procedures, and objective statistical criteria to determine their relative predictive performance. Within pharmaceutical market-demand prediction, comparative analysis is particularly important because different machine-learning algorithms employ distinct learning mechanisms, mathematical structures, optimization strategies, and computational assumptions that influence forecasting accuracy under varying market conditions. Pharmaceutical demand is characterized by nonlinear behavior resulting from interactions among epidemiological trends, prescription practices, pricing policies, promotional activities, healthcare utilization, reimbursement systems, demographic characteristics, seasonal disease patterns, regulatory changes, and competitive market dynamics (Ameer et al., 2019). These interacting variables generate complex forecasting environments in which no single analytical technique consistently provides optimal performance across every pharmaceutical product or healthcare setting. Comparative evaluation therefore enables researchers to examine algorithmic strengths and weaknesses under identical analytical conditions while identifying models that demonstrate superior robustness, predictive stability, computational efficiency, and forecasting accuracy. Such evaluation provides a scientifically rigorous basis for selecting appropriate machine-learning approaches according to specific pharmaceutical forecasting objectives rather than relying on generalized assumptions regarding algorithm superiority (Uddin et al., 2018).

The diversity of machine-learning methodologies further reinforces the importance of comparative investigation within pharmaceutical forecasting research. Linear regression provides interpretable baseline estimates under relatively simple relationships, whereas decision-tree algorithms identify hierarchical decision structures capable of modeling nonlinear interactions. Ensemble approaches such as random forests combine multiple decision trees to improve predictive stability and reduce variance across heterogeneous datasets. Gradient boosting techniques progressively improve forecasting performance by correcting residual prediction errors during successive learning iterations (Surdu & Győrgy, 2023). Support vector regression constructs optimized predictive boundaries capable of modeling nonlinear relationships using kernel functions, while artificial neural networks capture multidimensional interactions through interconnected computational layers. More advanced deep-learning architectures, including recurrent neural networks and long short-term memory models, analyze sequential temporal information by learning dependencies across historical observations, making them particularly suitable for time-series pharmaceutical forecasting. Extreme gradient boosting, LightGBM, CatBoost, adaptive boosting, and hybrid ensemble methods further expand forecasting capabilities by integrating multiple optimization strategies designed to improve predictive precision and computational efficiency (Azam et al., 2023). Because these algorithms differ substantially in feature selection mechanisms, computational complexity, scalability, interpretability, sensitivity to data quality, and resistance to overfitting, systematic comparison becomes essential for determining their suitability within diverse pharmaceutical forecasting scenarios.

Comparative machine-learning research additionally contributes to methodological standardization by promoting reproducible evaluation procedures applicable across different datasets, therapeutic markets, and healthcare systems. Objective comparison requires that algorithms be trained using identical preprocessing procedures, equivalent feature sets, consistent validation techniques, and uniform performance metrics so that differences in predictive outcomes accurately reflect algorithmic capability rather than variations in experimental design (Siddiqui et al., 2022). Standardized evaluation frameworks frequently include cross-validation, holdout testing, hyperparameter optimization, feature importance analysis, residual examination, sensitivity analysis, computational efficiency assessment, and statistical significance testing to establish comprehensive evidence regarding forecasting performance. Pharmaceutical organizations increasingly require transparent predictive systems capable of supporting strategic planning, inventory optimization, manufacturing coordination, commercial resource allocation, and market intelligence. Comparative quantitative studies therefore provide valuable empirical evidence by demonstrating how different machine-learning algorithms perform under realistic pharmaceutical market conditions while highlighting the circumstances in which particular forecasting methods exhibit stronger predictive capability (Fourkiotis & Tsadiras, 2024). Such systematic evaluation strengthens the scientific foundation of pharmaceutical sales forecasting research by emphasizing measurable analytical performance, objective statistical comparison, and evidence-based model selection across increasingly complex global healthcare markets.

The rapid advancement of machine-learning technologies has transformed predictive analytics across numerous industrial sectors, including finance, manufacturing, retail, transportation, healthcare, and pharmaceutical management. Within pharmaceutical industries, forecasting has traditionally relied on statistical time-series models, econometric techniques, expert judgment, and historical trend analysis to estimate medicine demand and commercial sales performance (Siddiqui et al., 2022). These conventional approaches have contributed significantly to pharmaceutical planning because they provide structured analytical frameworks for estimating future market behavior using historical observations. However, the increasing complexity of global pharmaceutical markets has introduced multidimensional relationships that are influenced simultaneously by epidemiological conditions, demographic transitions, prescription behavior, regulatory modifications, reimbursement mechanisms, promotional strategies, competitive product launches, healthcare utilization, pricing adjustments, supply-chain variability, and macroeconomic conditions. The interaction of these variables generates nonlinear demand structures that require advanced analytical techniques capable of identifying hidden relationships within extensive datasets (Patnaik et al., 2024). Machine-learning algorithms have therefore emerged as important forecasting tools because they can process large volumes of structured information while learning complex patterns that may not be represented adequately through conventional statistical assumptions. Although numerous predictive models have been introduced into pharmaceutical forecasting literature, considerable variation remains regarding the relative forecasting performance of different algorithms under comparable analytical conditions. Many investigations emphasize the effectiveness of a single predictive model without conducting systematic comparisons against alternative machine-learning approaches using identical datasets, standardized preprocessing methods, and objective evaluation metrics (Zingales et al., 2024). Such variation limits comprehensive understanding regarding which analytical methods consistently provide stronger forecasting performance across different pharmaceutical market environments.

Another important research gap concerns the diversity of datasets, forecasting horizons, evaluation procedures, and performance indicators applied across previous machine-learning investigations. Pharmaceutical forecasting studies have examined prescription demand, medicine utilization, inventory planning, hospital consumption, retail pharmacy sales, and commercial market performance using different sampling strategies, variable selections, feature engineering procedures, validation methods, and statistical performance measures. This methodological diversity creates challenges when attempting to compare findings across independent investigations because predictive outcomes may reflect differences in experimental design rather than algorithmic capability (Qu et al., 2020). Some studies evaluate forecasting models using relatively small datasets, whereas others employ large-scale commercial databases containing extensive temporal observations and multiple explanatory variables. Similarly, variations in data preprocessing, normalization, missing-value treatment, feature selection, hyperparameter optimization, cross-validation strategies, and testing procedures contribute additional inconsistency when interpreting predictive performance. Forecasting horizons also differ substantially across investigations, ranging from short-term daily demand estimation to monthly, quarterly, or annual pharmaceutical sales prediction. These methodological differences complicate objective assessment of machine-learning effectiveness because algorithm performance frequently depends upon data characteristics, forecasting intervals, feature complexity, and validation design (Bilal et al., 2024). Consequently, there remains a continuing need for quantitative research that evaluates multiple forecasting algorithms using standardized methodological procedures capable of supporting objective comparison across identical pharmaceutical datasets. Such comparative evaluation strengthens analytical consistency by reducing methodological variability while providing clearer evidence regarding relative forecasting capability.

The rationale for conducting a comparative quantitative analysis of machine-learning approaches therefore originates from the necessity to establish objective evidence regarding predictive performance within pharmaceutical sales forecasting and market-demand prediction (Ha et al., 2023). Pharmaceutical organizations increasingly depend upon accurate demand estimation to coordinate manufacturing schedules, procurement planning, inventory allocation, logistics management, commercial budgeting, and strategic market planning across highly competitive international environments. Decisions supported by inaccurate forecasts may influence medicine availability, warehouse utilization, production efficiency, operational expenditure, and organizational financial performance. Comparative evaluation enables forecasting algorithms to be assessed according to measurable statistical criteria while identifying differences in predictive accuracy, computational efficiency, robustness, generalization capability, and consistency under diverse commercial conditions (Qian et al., 2024). Such analysis provides empirical evidence concerning the suitability of individual machine-learning methods for pharmaceutical forecasting applications while strengthening methodological transparency through standardized experimental design. Quantitative comparison also contributes to scientific knowledge by distinguishing algorithmic strengths from dataset-specific influences, thereby improving understanding of how predictive models behave under realistic pharmaceutical market conditions (Bertolotti et al., 2024). The emphasis on objective statistical evaluation aligns with the broader development of evidence-based analytics within pharmaceutical management, where analytical decisions increasingly depend upon reproducible numerical evidence rather than subjective preference or isolated computational performance.

The present study was designed to conduct a comprehensive quantitative comparison of multiple machine-learning approaches for pharmaceutical sales forecasting and market-demand prediction using standardized analytical procedures and objective statistical evaluation (Fatima & Rahimi, 2024). The investigation focuses on the predictive capability of competing machine-learning algorithms in estimating pharmaceutical sales outcomes through numerical analysis of historical commercial and healthcare-related datasets. The quantitative design emphasizes measurable relationships between predictor variables and forecasting performance while maintaining methodological consistency throughout data preparation, model development, validation, and comparative assessment. Historical pharmaceutical sales information, together with relevant explanatory variables representing market conditions, commercial activities, demographic characteristics, healthcare utilization, seasonal influences, pricing behavior, and other operational indicators, forms the analytical basis for predictive model development (Jacob & Thomas, 2024). These structured numerical datasets provide opportunities to evaluate the capability of different machine-learning algorithms to recognize multidimensional relationships associated with pharmaceutical demand. The study adopts standardized preprocessing procedures, feature engineering techniques, training strategies, validation protocols, and statistical performance metrics to ensure that comparisons among forecasting models remain objective and scientifically reproducible. Such methodological consistency allows observed differences in forecasting performance to reflect algorithmic capability rather than inconsistencies in experimental design or data preparation (Safarishahrbijari, 2018).

A central characteristic of the present investigation is its emphasis on comparative evaluation rather than the assessment of an individual predictive model. Machine-learning algorithms differ substantially in mathematical formulation, optimization strategies, feature learning mechanisms, computational requirements, interpretability, scalability, and sensitivity to data characteristics. The study therefore evaluates multiple forecasting approaches under identical analytical conditions so that each algorithm receives equivalent opportunities to learn from the same pharmaceutical dataset (Borucka, 2023). Comparative performance is examined using established quantitative evaluation measures that assess prediction error, explanatory capability, forecasting precision, model stability, computational consistency, and overall predictive effectiveness. Standardized validation procedures ensure that each forecasting model is assessed according to equivalent statistical criteria, thereby improving analytical fairness and reducing methodological bias. The quantitative framework additionally supports examination of residual behavior, model robustness, generalization performance, and consistency across validation datasets. Such objective comparison contributes to a deeper understanding of the relative strengths and limitations associated with alternative machine-learning approaches when applied to pharmaceutical sales forecasting and market-demand prediction (Grobler-Dębska et al., 2024). The resulting analytical evidence provides a systematic basis for evaluating forecasting methodologies using transparent numerical criteria applicable across diverse pharmaceutical market environments.

The scope of the present study extends across the broader context of pharmaceutical business analytics by integrating predictive modeling, quantitative performance evaluation, and evidence-based comparative analysis within a unified methodological framework. Pharmaceutical forecasting represents an interdisciplinary analytical domain connecting healthcare management, business intelligence, operations research, statistics, data science, supply-chain management, and artificial intelligence (Feizabadi, 2022). The study contributes to this interdisciplinary perspective by emphasizing rigorous quantitative evaluation of forecasting algorithms using structured empirical procedures designed to support objective scientific investigation. Rather than relying upon descriptive interpretation or subjective algorithm selection, the research examines measurable forecasting outcomes through standardized statistical analysis capable of supporting reproducibility and analytical transparency. The methodological framework aligns with contemporary quantitative research practices that prioritize empirical validation, numerical evidence, and systematic comparison when evaluating predictive technologies. Through comprehensive comparative assessment of machine-learning approaches, the study establishes an organized analytical structure for examining pharmaceutical sales forecasting and market-demand prediction using consistent statistical methodology, objective performance measurement, and rigorous quantitative evaluation suitable for complex international pharmaceutical markets (Le et al., 2024).

The primary objective of this quantitative study was to conduct a comprehensive comparative analysis of machine-learning approaches for pharmaceutical sales forecasting and market-demand prediction by evaluating the predictive capability, accuracy, robustness, and overall analytical performance of multiple forecasting algorithms using standardized quantitative procedures. The study sought to investigate how different machine-learning models perform when trained and validated on pharmaceutical sales and market-demand datasets containing multidimensional variables associated with historical sales performance, pricing behavior, promotional activities, inventory characteristics, seasonal fluctuations, demographic factors, healthcare utilization, prescription trends, and market conditions. Particular emphasis was placed on identifying differences in forecasting accuracy among competing algorithms through objective statistical evaluation using established performance metrics and validation techniques. The study further aimed to determine the extent to which machine-learning algorithms could effectively capture nonlinear relationships, temporal dependencies, and complex interactions that characterize pharmaceutical demand patterns across diverse market environments.

Another objective was to examine the consistency, generalization capability, computational stability, and prediction reliability of alternative forecasting approaches under identical preprocessing procedures, feature engineering methods, training strategies, and testing protocols to ensure fair and unbiased comparison. The investigation also sought to evaluate model performance using standardized quantitative indicators that measure prediction error, explanatory power, forecasting precision, and overall predictive effectiveness while identifying the relative strengths and limitations of each machine-learning technique within pharmaceutical forecasting applications. Furthermore, the study intended to establish an evidence-based analytical framework for comparing predictive models by integrating systematic data preparation, statistical validation, cross-validation procedures, and objective performance assessment into a unified quantitative methodology. The research additionally aimed to contribute to pharmaceutical business analytics by generating empirical evidence regarding the suitability of different machine-learning approaches for supporting sales forecasting, inventory planning, production scheduling, demand estimation, supply-chain coordination, commercial decision-making, and strategic resource allocation. Through rigorous comparative quantitative analysis, the study sought to provide a scientifically grounded assessment of forecasting methodologies capable of improving analytical accuracy and strengthening data-driven decision-making within increasingly complex pharmaceutical markets while maintaining methodological transparency, reproducibility, and statistical rigor throughout the entire research process.

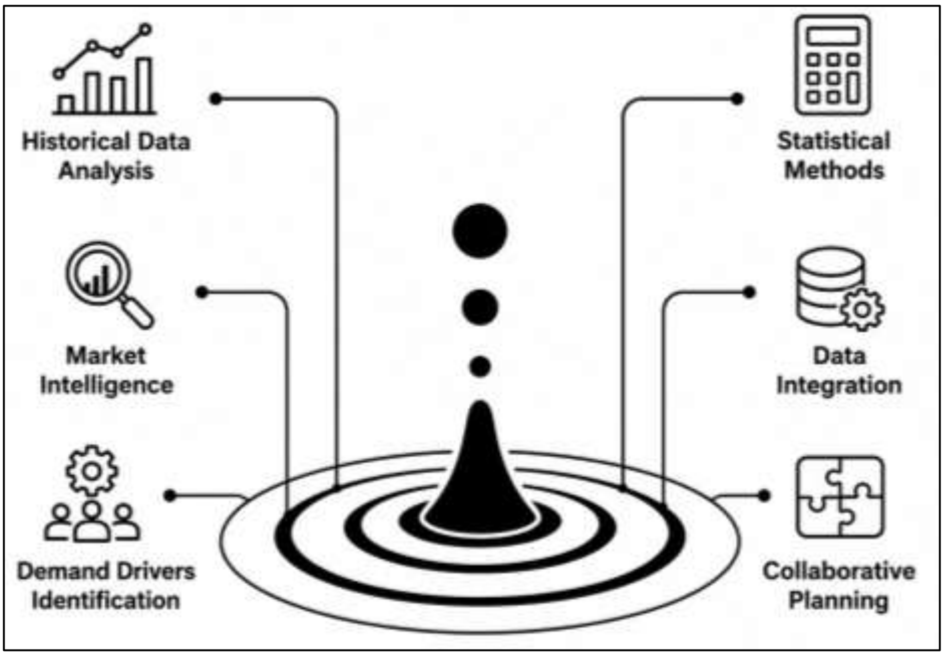
LITERATURE REVIEW

The literature review establishes the theoretical, conceptual, methodological, and empirical foundation for understanding the application of machine-learning approaches in pharmaceutical sales forecasting and market-demand prediction. Pharmaceutical forecasting has evolved from conventional statistical estimation toward data-driven predictive analytics because pharmaceutical markets generate extensive multidimensional datasets characterized by nonlinear relationships, seasonal variability, heterogeneous customer behavior, healthcare utilization dynamics, pricing fluctuations, promotional activities, regulatory influences, epidemiological changes, and complex supply-chain interactions (Armstrong & Green, 2018; Md. Abdur & Iftekhar, 2021). Machine-learning algorithms have increasingly become central analytical tools for extracting meaningful predictive information from these complex datasets, allowing organizations to improve forecasting accuracy and operational decision-making. The existing body of literature demonstrates substantial growth in the application of supervised learning, ensemble learning, deep learning, and hybrid artificial intelligence techniques for forecasting commercial pharmaceutical performance. However, previous studies differ considerably in algorithm selection, predictor variables, validation procedures, forecasting horizons, dataset characteristics, and statistical evaluation methods. Consequently, a comprehensive synthesis of quantitative evidence is required to understand the relative strengths, methodological developments, and predictive capabilities of competing machine-learning approaches within pharmaceutical forecasting environments (Kazi Rakib Hasan & Uddin, 2022; Syam & Sharma, 2018). This literature review therefore critically examines theoretical perspectives, forecasting variables, quantitative measurement approaches, machine-learning algorithms, model evaluation techniques, empirical findings, and methodological developments that collectively support comparative quantitative analysis. The review is organized systematically from conceptual foundations toward advanced predictive modeling, providing an integrated understanding of pharmaceutical sales forecasting while identifying the quantitative evidence necessary to support the present comparative investigation.

Pharmaceutical Sales Forecasting and Market-Demand Prediction

Pharmaceutical sales forecasting and market-demand prediction represent two closely connected quantitative concepts within pharmaceutical business analytics, supply-chain planning, and healthcare market management. Pharmaceutical sales forecasting refers to the systematic estimation of future sales volume, revenue movement, prescription demand, product uptake, and distribution requirements by analyzing historical sales records, market behavior, therapeutic category performance, pricing patterns, promotional activity, physician prescribing trends, and healthcare utilization indicators (Mohiul & Badrul, 2022; Siddiqui et al., 2022). Earlier literature described sales forecasting as an evidence-based planning process that supports production scheduling, inventory control, financial planning, procurement decisions, and commercial resource allocation. In pharmaceutical industries, forecasting carries greater complexity than ordinary retail sales estimation because medicine demand is shaped by clinical need, disease prevalence, patient demographics, insurance coverage, regulatory approval, reimbursement rules, physician decision-making, seasonal illness patterns, and product life-cycle stages. Market-demand prediction, on the other hand, refers to the broader analytical process of estimating how much demand may exist for a pharmaceutical product across specific markets, regions, patient groups, healthcare institutions, or therapeutic areas (Kumar et al., 2024; Mukut Kanti & Rony, 2022). It includes not only projected sales transactions but also expected patient need, prescription potential, market penetration, treatment adoption, competitive pressure, and consumption behavior. Prior quantitative studies on pharmaceutical analytics, healthcare demand modeling, retail forecasting, supply-chain forecasting, artificial intelligence, time-series analysis, and machine-learning prediction collectively showed that forecasting and demand prediction depend on structured numerical evidence rather than subjective judgment. The literature also indicated that accurate prediction requires integration of historical trends, external market indicators, operational data, and healthcare system variables (Kallivokas et al., 2024; Sadia et al., 2022). Therefore, pharmaceutical sales forecasting can be understood as a sales-centered predictive activity, while pharmaceutical market-demand prediction represents a wider demand-centered analytical process that explains the measurable forces influencing product utilization and market behavior.

Figure 3: Pharmaceutical Demand Forecasting Framework



Pharmaceutical demand systems possess distinctive characteristics that differentiate them from demand systems observed in most consumer goods industries because medicine utilization is influenced by a combination of clinical, economic, demographic, regulatory, and operational factors rather than consumer preference alone (Debasish, 2021; Mukut Kanti, 2020). The literature consistently described pharmaceutical demand as dynamic, multidimensional, and highly sensitive to changes in disease prevalence, healthcare accessibility, physician prescribing practices, patient adherence, reimbursement policies, product availability, generic competition, and government regulation (Papanagnou & Matthews-Amune, 2018). Unlike conventional retail markets, pharmaceutical purchasing frequently involves multiple stakeholders, including physicians, pharmacists, hospitals, insurance providers, healthcare administrators, regulatory agencies, distributors, and patients, each contributing to the overall demand process. Previous quantitative investigations indicated that pharmaceutical demand demonstrates considerable temporal variation resulting from seasonal disease outbreaks, chronic disease prevalence, vaccination campaigns, demographic transitions, public health interventions, and epidemiological changes (Abdur & Iftekhar, 2021; Chowdhury & Tonay, 2022). Demand systems also exhibit regional heterogeneity because healthcare infrastructure, economic development, population characteristics, insurance coverage, prescribing guidelines, and medicine accessibility differ substantially across countries and healthcare systems (Alzoubi et al., 2024; Md Tohidul, 2023). Earlier studies further emphasized that pharmaceutical demand often contains nonlinear relationships, delayed market responses, irregular purchasing cycles, product substitution effects, and interactions among multiple explanatory variables, making prediction considerably more complex than traditional commercial forecasting (Badrul & Mominul, 2023; Zahidul & Begum, 2022; Nishat & Mst Kaniz, 2022). Literature on healthcare analytics additionally highlighted the growing influence of digital health records, electronic prescribing systems, pharmacy management platforms, and integrated healthcare databases, which have expanded the availability of structured information describing medicine utilization (Badrul & Mominul, 2023; Tomar & Sharma, 2024). Collectively, previous studies portrayed pharmaceutical demand systems as complex analytical environments requiring quantitative approaches capable of processing extensive multidimensional datasets while accurately representing the interactions among healthcare, commercial, and operational variables that influence medicine consumption.

The commercial importance of forecasting within pharmaceutical industries has been widely emphasized throughout the literature because forecasting directly supports strategic planning, operational efficiency, financial management, inventory optimization, production scheduling, procurement coordination, distribution planning, and market competitiveness. Earlier research consistently demonstrated that reliable sales forecasts enable pharmaceutical manufacturers to balance production capacity with anticipated market demand while minimizing inventory shortages, excess stock, product expiration, warehouse congestion, and unnecessary operating costs (Hossan & Adar, 2023; Sazvar et al., 2022). The literature further indicated that accurate forecasting contributes to efficient procurement of raw materials, optimization of manufacturing schedules, effective allocation of logistics resources, and improved coordination among suppliers, distributors, wholesalers, hospitals, and retail pharmacies. Commercial forecasting also supports pricing decisions, promotional planning, sales force deployment, budgeting, revenue estimation, and investment evaluation by providing quantitative evidence regarding expected market performance. Previous investigations showed that forecasting plays a particularly important role during product launches, patent expiration periods, generic medicine competition, therapeutic market expansion, and changes in reimbursement or regulatory policies because these events frequently alter purchasing behavior and sales performance (Mitra et al., 2024; Risha & Khalid, 2023). Studies examining pharmaceutical supply chains additionally reported that improved forecasting enhances medicine availability, reduces distribution inefficiencies, strengthens service levels, and supports continuity of patient care through better inventory management across healthcare networks. The literature also recognized forecasting as an essential component of pharmaceutical business intelligence because predictive analytics transforms large volumes of commercial and healthcare data into measurable information supporting evidence-based managerial decision-making. Overall, previous research consistently identified pharmaceutical forecasting as a strategic commercial function that integrates quantitative analysis with operational planning to improve organizational performance, strengthen supply-chain coordination, optimize financial outcomes, and enhance the efficiency of medicine distribution within increasingly complex global pharmaceutical markets (Harshavardhini et al., 2024).

The global pharmaceutical market operates within a highly interconnected commercial and healthcare environment in which medicine demand is influenced by demographic transitions, disease epidemiology, healthcare expenditure, technological innovation, regulatory frameworks, and international supply-chain activities. The literature consistently described the pharmaceutical industry as one of the most data-intensive sectors because commercial performance depends on continuous interactions among manufacturers, regulatory agencies, healthcare providers, wholesalers, pharmacies, insurers, and patients across multiple geographic regions (Crawford et al., 2020). Earlier studies indicated that global pharmaceutical markets exhibit substantial variation in therapeutic demand due to differences in population aging, chronic disease prevalence, healthcare infrastructure, national reimbursement systems, income levels, prescribing guidelines, and access to essential medicines. These structural differences generate heterogeneous market behavior that requires forecasting models capable of accommodating regional diversity while maintaining predictive consistency. Previous investigations also emphasized that pharmaceutical commercialization extends beyond product manufacturing because market performance is shaped by regulatory approvals, patent protection, generic competition, biosimilar introduction, pricing regulations, product availability, distribution efficiency, and commercial marketing activities (Qureshi et al., 2024). Literature examining pharmaceutical business analytics further highlighted that globalization has increased operational complexity by extending supply networks across multiple countries, thereby making demand estimation increasingly dependent upon accurate integration of international commercial and healthcare data. Researchers consistently recognized that digital transformation, enterprise information systems, pharmacy databases, electronic prescribing platforms, and healthcare information technologies have expanded the availability of structured data describing medicine utilization across international markets (Liu et al., 2022). Consequently, pharmaceutical forecasting has become increasingly dependent upon advanced quantitative analytics capable of processing multidimensional information generated from diverse commercial and healthcare environments.

The literature further identified numerous determinants that collectively influence pharmaceutical sales performance, demonstrating that medicine sales cannot be explained by a single commercial factor. Historical sales volume remains one of the strongest predictors because previous purchasing behavior frequently reflects recurring demand patterns and established prescribing practices. However, earlier studies emphasized that sales performance is simultaneously affected by product pricing strategies, promotional expenditure, physician prescribing preferences, patient adherence, disease prevalence, healthcare utilization, reimbursement policies, pharmacy inventory availability, distribution efficiency, competitor activities, product life-cycle stage, seasonal illness patterns, and macroeconomic conditions (Kulkov, 2021). Research also indicated that therapeutic category characteristics influence market performance because chronic disease medications often demonstrate relatively stable purchasing behavior, whereas products associated with acute illnesses display stronger seasonal fluctuations. Additional investigations highlighted the importance of demographic variables, including population growth, age distribution, urbanization, and socioeconomic conditions, in explaining medicine consumption across different healthcare systems. Literature on pharmaceutical marketing demonstrated that physician engagement, educational campaigns, digital promotion, brand recognition, and market access initiatives contribute significantly to commercial performance, while supply-chain efficiency influences medicine availability and customer satisfaction (Jian et al., 2020). Collectively, previous studies portrayed pharmaceutical sales performance as the result of complex interactions among clinical, operational, economic, demographic, and commercial variables, reinforcing the necessity of quantitative forecasting approaches capable of simultaneously analyzing multiple determinants within highly dynamic pharmaceutical markets.

Demand uncertainty has consistently been identified in the literature as one of the most significant analytical challenges affecting pharmaceutical forecasting because medicine demand is subject to continuous variation arising from clinical, commercial, operational, and regulatory influences (Bilal et al., 2024). Earlier research described pharmaceutical demand uncertainty as the inability to estimate future medicine utilization with complete precision due to fluctuations in disease outbreaks, prescription behavior, healthcare utilization, treatment guidelines, reimbursement changes, product substitution, inventory adjustments, and patient purchasing patterns. Unlike many consumer products, pharmaceutical demand is frequently affected by unexpected epidemiological events, regulatory interventions, product recalls, medicine shortages, public health campaigns, and changes in healthcare policy, all of which may substantially alter purchasing behavior within relatively short periods. Previous studies further explained that uncertainty is amplified by the extensive number of interacting variables influencing pharmaceutical markets, including demographic transitions, physician prescribing decisions, patient adherence, seasonal disease incidence, competitor product launches, pricing modifications, and distribution network performance (Merkuryeva et al., 2019). The literature also emphasized that uncertainty exists across multiple forecasting horizons because short-term demand may respond to temporary operational factors, whereas long-term demand reflects broader demographic, clinical, and market developments. Researchers consistently reported that forecasting performance deteriorates when analytical models fail to capture nonlinear interactions, temporal variability, hidden demand patterns, and irregular market behavior (Subramanian, 2021). Consequently, pharmaceutical demand uncertainty has been widely recognized as a multidimensional phenomenon requiring sophisticated quantitative methodologies capable of processing large and heterogeneous datasets while adapting to changing commercial and healthcare conditions.

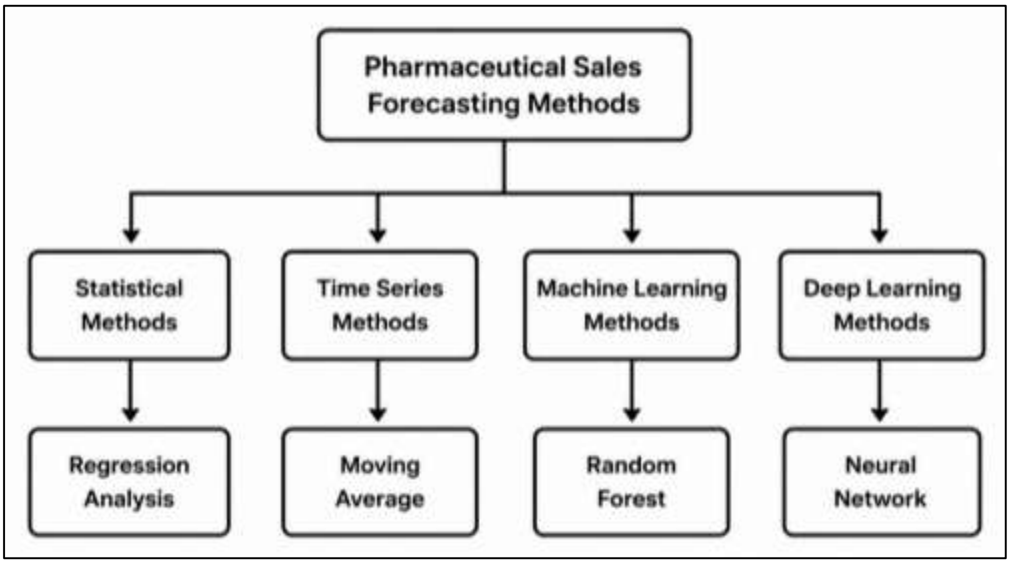
The literature also documented a clear evolution in forecasting methodology from conventional statistical approaches toward machine-learning-based predictive analytics. Earlier forecasting studies primarily relied on statistical techniques that estimated future sales using historical observations, trend analysis, seasonal decomposition, regression modeling, and time-series methods (Nguyen et al., 2023). These approaches established an important analytical foundation for pharmaceutical forecasting because they introduced systematic procedures for measuring historical market behavior and estimating future demand using structured numerical evidence. As pharmaceutical datasets expanded in size and complexity, researchers increasingly recognized that traditional statistical methods encountered limitations when analyzing nonlinear relationships, high-dimensional variables, heterogeneous datasets, and rapidly changing market conditions. This recognition contributed to the gradual adoption of machine-learning techniques capable of learning complex relationships directly from historical data without relying exclusively on predetermined statistical assumptions (Siddiqui et al., 2022). The literature demonstrated that decision trees, random forests, support vector machines, artificial neural networks, gradient boosting algorithms, deep-learning architectures, and ensemble learning methods progressively became prominent forecasting tools because of their ability to process extensive commercial and healthcare information while improving predictive accuracy across complex datasets. Previous comparative investigations consistently reported that machine-learning approaches enhanced the analysis of multidimensional pharmaceutical data by identifying hidden interactions among sales, demographic, operational, epidemiological, and commercial variables (Amalnick et al., 2020). Collectively, the literature portrayed this methodological transition as an important advancement in pharmaceutical business analytics, reflecting the increasing integration of artificial intelligence, predictive modeling, and quantitative data analysis into modern pharmaceutical sales forecasting and market-demand prediction.

Quantitative Theoretical Foundations Supporting Machine-Learning Forecasting

Predictive Analytics Theory and Statistical Learning Theory collectively provide the principal quantitative foundation for machine-learning forecasting by explaining how historical observations can be transformed into reliable predictions of future outcomes through systematic analysis of numerical data. Predictive Analytics Theory emerged from the integration of statistics, operations research, data science, and business intelligence, emphasizing that historical information contains measurable patterns capable of supporting evidence-based forecasting and decision-making (Nguyen et al., 2022). Within pharmaceutical industries, this theoretical perspective views sales transactions, prescription records, inventory movements, pricing behavior, healthcare utilization, promotional activities, demographic information, epidemiological conditions, and market indicators as quantitative variables that collectively explain future commercial performance. Earlier literature consistently described predictive analytics as a structured process involving data acquisition, preprocessing, feature extraction, model development, validation, and performance evaluation, where forecasting accuracy depends upon the quality of historical data and the analytical capability of predictive models. Pharmaceutical demand is influenced by numerous interconnected variables that rarely operate independently, making predictive analytics particularly valuable because it integrates heterogeneous information into unified forecasting frameworks (Yani & Aamer, 2023). Previous studies further emphasized that predictive analytics supports commercial planning, inventory optimization, production scheduling, revenue estimation, procurement management, and strategic decision-making by transforming large datasets into measurable forecasting outputs. The literature also highlighted that predictive analytics contributes to objective managerial decision-making because forecasting performance is evaluated through standardized quantitative indicators rather than subjective business judgment. Consequently, Predictive Analytics Theory provides an important conceptual foundation for comparative machine-learning research by establishing that reliable pharmaceutical forecasting depends upon systematic extraction of meaningful predictive information from historical commercial and healthcare datasets (Festa et al., 2022).

Statistical Learning Theory complements Predictive Analytics Theory by providing the quantitative principles that explain how computational models learn relationships between explanatory variables and target outcomes while minimizing prediction error and improving generalization across unseen observations. Earlier theoretical literature described statistical learning as the process through which algorithms identify hidden structures, patterns, and dependencies within numerical datasets by learning from historical examples instead of relying exclusively on predetermined mathematical assumptions (Viboud et al., 2018). The theory emphasizes that predictive performance depends not only on fitting historical observations but also on the ability of analytical models to produce consistent and reliable predictions when applied to new data representing similar market conditions. Previous investigations indicated that pharmaceutical datasets frequently contain multidimensional variables, nonlinear interactions, temporal variation, heterogeneous observations, and substantial market complexity, making statistical learning especially relevant for commercial forecasting applications. Statistical Learning Theory also introduced the concept of balancing model complexity with predictive stability so that forecasting algorithms avoid excessive adaptation to historical noise while maintaining strong generalization capability (Masini et al., 2023). The literature consistently demonstrated that machine-learning algorithms including regression models, decision trees, support vector machines, ensemble techniques, and neural networks all operate according to statistical learning principles because each algorithm estimates predictive relationships from observed data through systematic optimization procedures.

Figure 4: Pharmaceutical Forecasting Methods Framework



Researchers further argued that statistical learning provides an objective framework for evaluating forecasting performance using standardized validation methods, comparative statistical analysis, and quantitative accuracy measures, thereby supporting transparent comparison among competing predictive models (Sonkavde et al., 2023). Collectively, Predictive Analytics Theory and Statistical Learning Theory establish a rigorous quantitative foundation for pharmaceutical sales forecasting by explaining how machine-learning algorithms transform historical pharmaceutical information into measurable predictions capable of supporting evidence-based commercial planning and market-demand estimation.

Computational Intelligence Theory provides an important quantitative foundation for machine-learning forecasting by explaining how intelligent computational systems analyze complex datasets, recognize hidden relationships, adapt to changing information, and generate reliable predictions from historical observations (Abolghasemi, 2023). The literature consistently described computational intelligence as a multidisciplinary field integrating artificial intelligence, machine learning, neural computation, evolutionary computation, fuzzy systems, and adaptive learning to solve analytical problems characterized by uncertainty, multidimensionality, and nonlinear relationships. Within pharmaceutical sales forecasting and market-demand prediction, Computational Intelligence Theory views commercial and healthcare data as interconnected sources of information that can be analyzed to reveal meaningful patterns influencing medicine demand and sales performance. Previous studies emphasized that pharmaceutical markets generate extensive datasets containing prescription records, sales transactions, inventory movements, promotional activities, pricing behavior, demographic characteristics, epidemiological indicators, healthcare utilization records, and supply-chain information, all of which interact simultaneously to influence commercial outcomes (Chou et al., 2023). Conventional analytical approaches often encounter challenges when processing these heterogeneous relationships because pharmaceutical demand rarely follows simple linear patterns. The literature therefore recognized computational intelligence as an effective theoretical framework capable of supporting adaptive learning from highly complex datasets while improving predictive accuracy under changing market conditions. Earlier investigations further highlighted that computational intelligence enables forecasting systems to identify subtle interactions among explanatory variables without requiring rigid analytical assumptions regarding variable relationships (Camporeale, 2019). This adaptability allows predictive models to represent real-world pharmaceutical environments more effectively than many traditional forecasting approaches. Researchers also noted that computational intelligence contributes to objective quantitative analysis through continuous model refinement, pattern discovery, predictive optimization, and systematic evaluation of forecasting performance using standardized numerical criteria. Overall, Computational Intelligence Theory establishes that machine-learning algorithms function as adaptive analytical systems capable of learning complex commercial behavior from historical pharmaceutical data while improving the reliability of demand prediction and sales forecasting within dynamic healthcare markets.

Data Mining Theory complements Computational Intelligence Theory by explaining the systematic extraction of meaningful knowledge, hidden structures, and predictive information from large collections of structured and unstructured data (Barrera-Animas et al., 2022). Earlier literature consistently described data mining as a knowledge discovery process involving data preparation, cleaning, transformation, pattern identification, relationship discovery, predictive modeling, and interpretation of analytical findings. Within pharmaceutical industries, Data Mining Theory provides the conceptual basis for transforming extensive healthcare and commercial databases into actionable forecasting information that supports evidence-based decision-making. Previous studies reported that pharmaceutical organizations accumulate large volumes of transactional and operational information through enterprise resource planning systems, pharmacy management platforms, electronic prescribing systems, hospital databases, insurance claims, inventory records, customer purchasing histories, and digital marketing platforms (Zeineddine et al., 2021). These datasets contain valuable information regarding medicine utilization, sales performance, prescribing behavior, market segmentation, seasonal demand variation, and consumer purchasing patterns that cannot be identified effectively through manual inspection alone. The literature demonstrated that data mining techniques systematically identify recurring demand patterns, customer behavior, product associations, temporal trends, and market relationships that strengthen forecasting accuracy and improve analytical understanding of pharmaceutical demand systems. Researchers further emphasized that effective data mining depends upon data quality, feature relevance, preprocessing consistency, and systematic validation because inaccurate or incomplete information may reduce predictive performance (Januschowski et al., 2020). Previous investigations also highlighted that data mining supports machine-learning forecasting by providing structured datasets, informative predictor variables, and meaningful feature representations that improve algorithm learning and comparative evaluation. Collectively, Computational Intelligence Theory and Data Mining Theory establish complementary quantitative foundations for pharmaceutical forecasting by demonstrating how adaptive computational methods and systematic knowledge discovery transform extensive pharmaceutical datasets into reliable predictive information capable of supporting objective forecasting, commercial planning, and market-demand prediction (Badillo et al., 2020).

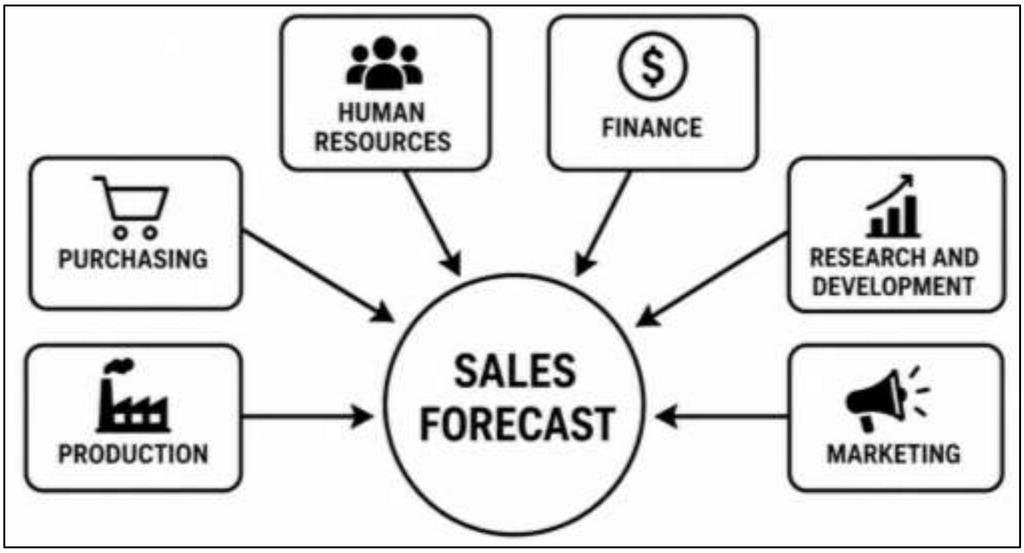
Pattern Recognition Theory provides a significant quantitative foundation for machine-learning forecasting because it explains how computational systems identify recurring structures, classify observations, and distinguish meaningful relationships from large collections of historical data. The literature consistently described pattern recognition as the analytical process of detecting regularities, similarities, and hidden associations within multidimensional datasets to support accurate prediction and classification (Papacharalampous et al., 2019). In pharmaceutical sales forecasting and market-demand prediction, this theoretical perspective assumes that historical sales records, prescription volumes, inventory transactions, pricing behavior, promotional activities, disease prevalence, healthcare utilization, seasonal demand, demographic characteristics, and commercial indicators contain identifiable patterns that can be systematically learned and applied to future forecasting tasks. Previous studies emphasized that pharmaceutical markets exhibit recurring demand cycles, regional purchasing behavior, therapeutic consumption trends, and customer response patterns that become observable when large datasets are analyzed using advanced computational techniques. The literature further indicated that effective forecasting depends on the ability of predictive models to recognize both stable and changing demand structures while distinguishing meaningful information from random variation (Syam & Sharma, 2018). Earlier investigations demonstrated that machine-learning algorithms such as decision trees, support vector machines, neural networks, ensemble learning methods, and deep-learning architectures rely extensively on pattern recognition principles because these models identify complex relationships among predictor variables without depending solely on predefined statistical assumptions. Researchers also argued that pharmaceutical datasets frequently contain heterogeneous variables originating from healthcare systems, commercial operations, logistics, epidemiological surveillance, and market intelligence, making pattern recognition particularly valuable for identifying interactions that influence medicine demand (Liu et al., 2020). Previous comparative studies reported that successful recognition of temporal, behavioral, and commercial patterns contributes directly to improved forecasting accuracy, enhanced market segmentation, stronger inventory planning, and more reliable prediction of pharmaceutical sales performance. Collectively, the literature established Pattern Recognition Theory as an essential conceptual framework supporting machine-learning forecasting by demonstrating that predictive accuracy depends upon systematic identification of recurring structures embedded within historical pharmaceutical and healthcare data (Zhuo et al., 2018).

Artificial Intelligence Decision Theory complements Pattern Recognition Theory by explaining how intelligent computational systems convert analytical knowledge into objective decision support through automated reasoning, predictive inference, and evidence-based evaluation. Earlier literature characterized Artificial Intelligence Decision Theory as an interdisciplinary framework integrating artificial intelligence, decision sciences, computational reasoning, optimization methods, and predictive analytics to improve decision quality under conditions of uncertainty and complexity. Within pharmaceutical industries, the theory views forecasting as a decision-support process in which analytical models assist managers by generating reliable estimates of future sales, medicine demand, inventory requirements, production schedules, and commercial resource allocation (Nikou et al., 2019). Previous studies emphasized that pharmaceutical decision-making involves simultaneous consideration of numerous quantitative variables, including sales history, prescription behavior, pricing policies, promotional investment, healthcare utilization, reimbursement systems, demographic characteristics, regulatory changes, and supply-chain performance. The literature consistently indicated that traditional decision-making approaches relying primarily on human judgment may encounter limitations when processing high-dimensional datasets characterized by nonlinear interactions and rapidly changing market conditions. Artificial Intelligence Decision Theory therefore supports the application of machine-learning algorithms capable of evaluating multiple variables simultaneously while identifying optimal predictive relationships from historical evidence (Schmidt et al., 2017). Earlier investigations demonstrated that intelligent forecasting systems improve consistency, transparency, and analytical objectivity by reducing dependence on subjective interpretation and replacing it with systematic quantitative evaluation. Researchers further highlighted that artificial intelligence supports adaptive commercial planning by continuously learning from new pharmaceutical data, refining predictive capability, and improving forecasting performance through repeated analytical optimization. Comparative empirical studies also reported that artificial intelligence–based decision frameworks strengthen pharmaceutical business intelligence, inventory management, supply-chain coordination, demand planning, and strategic commercial analysis by providing measurable evidence supporting managerial decisions (Schweidtmann et al., 2021). Overall, Pattern Recognition Theory and Artificial Intelligence Decision Theory collectively establish a comprehensive quantitative foundation for machine-learning forecasting by explaining how computational systems recognize meaningful demand patterns and transform analytical knowledge into objective, data-driven decision support within complex pharmaceutical markets.

Quantitative Variables Influencing Pharmaceutical Sales Forecasting

Historical sales volume, prescription frequency, drug pricing dynamics, and promotional and marketing expenditure have consistently been identified in the literature as the most influential quantitative variables affecting pharmaceutical sales forecasting and market-demand prediction (Wei et al., 2018). Earlier studies described historical sales volume as the principal indicator of future commercial performance because it captures recurring purchasing behavior, long-term market trends, product adoption rates, seasonal fluctuations, and customer loyalty across different therapeutic categories. Historical transaction records generated through pharmaceutical manufacturers, wholesalers, hospitals, and retail pharmacies provide extensive numerical information that enables forecasting models to identify stable demand patterns while recognizing changes associated with product maturity, competitive activity, and healthcare utilization. Researchers consistently emphasized that historical sales data serve as the foundation for predictive modeling because they reflect cumulative interactions among commercial, operational, demographic, and clinical variables (Rundo et al., 2019).

Figure 5: Pharmaceutical Sales Forecasting Drivers



Alongside historical sales, prescription frequency has been recognized as another critical indicator because prescription records measure actual medicine utilization within healthcare systems rather than only commercial transactions. The literature demonstrated that prescription frequency reflects physician treatment decisions, disease prevalence, patient adherence, healthcare accessibility, clinical guidelines, and therapeutic effectiveness, thereby providing direct evidence of underlying pharmaceutical demand. Previous investigations also reported that combining prescription information with historical sales improves forecasting performance because prescriptions frequently indicate future purchasing behavior before complete sales data become available. Collectively, these variables provide complementary information regarding both commercial activity and healthcare consumption, making them essential predictors within quantitative forecasting models designed to estimate pharmaceutical sales accurately across diverse healthcare environments (Merkuryeva et al., 2019).

Drug pricing dynamics and promotional and marketing expenditure further strengthen pharmaceutical forecasting because they directly influence purchasing behavior, product accessibility, physician awareness, and market competitiveness. Earlier literature consistently reported that medicine pricing affects patient affordability, reimbursement eligibility, procurement decisions, product substitution, and competitive positioning, all of which contribute to measurable variation in pharmaceutical demand. Pricing strategies differ substantially across branded medicines, generic pharmaceuticals, biosimilars, and specialty therapeutics, resulting in diverse commercial responses that forecasting models must capture accurately (Siddiqui et al., 2022). Researchers observed that pricing adjustments frequently alter purchasing volume, particularly in healthcare systems where patients bear a significant proportion of medicine costs or where reimbursement policies are sensitive to price changes. Promotional and marketing expenditure has also been widely recognized as a major commercial determinant because pharmaceutical organizations invest considerable resources in physician detailing, scientific meetings, educational programs, digital promotion, product advertising, and sales-force activities to increase product awareness and market penetration. Previous empirical investigations consistently demonstrated positive associations between marketing investment and pharmaceutical sales performance, particularly during product introduction, therapeutic expansion, and periods of intense market competition. The literature further emphasized that promotional effectiveness depends upon therapeutic category, physician engagement, product differentiation, and healthcare regulations governing pharmaceutical communication (Subramanian, 2021). Studies examining machine-learning forecasting consistently incorporated pricing information and promotional expenditure alongside historical sales and prescription frequency because these variables improve the ability of predictive models to explain fluctuations in medicine demand under changing commercial conditions. Collectively, the literature established that historical sales volume, prescription frequency, pricing behavior, and promotional investment represent interconnected quantitative variables that provide a comprehensive explanation of pharmaceutical market behavior while significantly enhancing the predictive capability of forecasting models used in pharmaceutical business analytics (Yani & Aamer, 2023).

Physician prescribing behavior, pharmacy inventory levels, and seasonal disease patterns have been widely emphasized in the literature as interconnected quantitative variables that strongly influence pharmaceutical sales forecasting and market-demand prediction. Physician prescribing behavior represents one of the most direct determinants of pharmaceutical demand because physicians act as clinical decision-makers who determine medicine selection, treatment duration, dosage patterns, therapeutic substitution, and patient access to prescription products. Earlier studies described prescribing behavior as a measurable outcome shaped by clinical guidelines, diagnostic practices, physician experience, patient characteristics, drug efficacy evidence, safety profile, brand familiarity, reimbursement coverage, and pharmaceutical promotion (Ali et al., 2024). Variations in prescribing behavior across hospitals, clinics, specialties, regions, and healthcare systems create measurable differences in product demand and sales volume. Literature on pharmaceutical forecasting consistently indicated that prescription trends provide valuable signals for estimating future sales because increased prescribing frequency usually precedes or accompanies higher medicine consumption. Pharmacy inventory levels also serve as important forecasting variables because they reflect product availability, replenishment cycles, stock turnover, purchasing expectations, and operational responses to anticipated demand. Previous studies demonstrated that inadequate inventory may suppress observed sales by creating stockouts, whereas excessive inventory may indicate overestimation of demand, slow-moving products, or inefficient procurement planning (Fourkiotis & Tsadiras, 2024). Inventory records from hospitals, wholesalers, distributors, and retail pharmacies therefore provide quantitative evidence about the relationship between expected demand and actual supply-chain behavior. Seasonal disease patterns further complicate forecasting because medicine utilization often changes according to recurring epidemiological cycles, climatic conditions, infection rates, allergy seasons, and public health events. Earlier literature reported that respiratory medicines, antibiotics, antipyretics, vaccines, allergy medications, and antiviral products frequently show seasonal variation in sales and prescription volume. Machine-learning forecasting studies have therefore incorporated prescribing indicators, inventory measures, and seasonal disease variables to improve prediction accuracy by capturing clinical demand, operational availability, and temporal fluctuations simultaneously (Merkuryeva et al., 2019). Collectively, these studies established that pharmaceutical sales forecasting becomes more reliable when predictive models integrate healthcare decision behavior, supply-chain inventory signals, and seasonally recurring disease patterns within a unified quantitative framework.

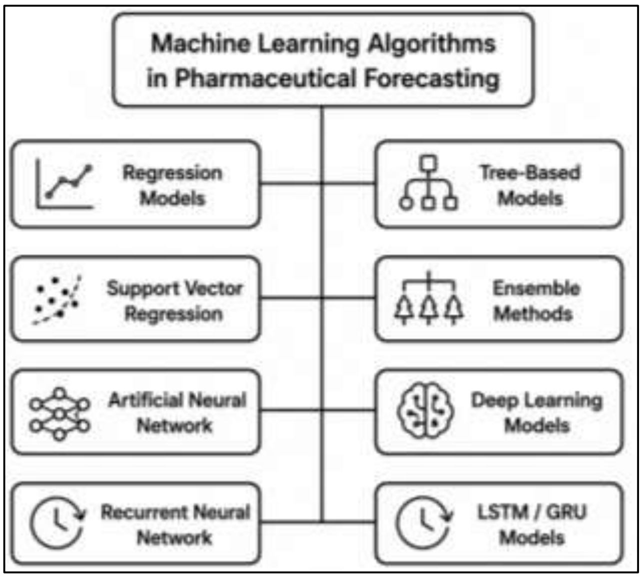
Patient demographics, healthcare utilization indicators, insurance reimbursement policies, and competitor product activities have consistently been recognized in the literature as important quantitative variables influencing pharmaceutical sales forecasting and market-demand prediction because they collectively explain variations in medicine consumption, purchasing behavior, and commercial market performance. Patient demographic characteristics provide measurable information regarding population size, age distribution, gender composition, socioeconomic status, educational attainment, geographic location, disease prevalence, and health-related behaviors, all of which contribute to differences in pharmaceutical demand across regions and healthcare systems (Siddiqui et al., 2022). Earlier studies demonstrated that demographic transitions, particularly population aging and the increasing prevalence of chronic diseases, significantly influence long-term medicine utilization because older populations generally require more frequent pharmacological treatment than younger age groups. The literature also reported that urbanization, population density, household income, and educational levels affect healthcare accessibility, treatment adherence, and medicine purchasing patterns. Healthcare utilization indicators further strengthen forecasting models because they measure the frequency of interactions between patients and healthcare providers through hospital admissions, outpatient consultations, emergency department visits, diagnostic procedures, specialist referrals, and treatment episodes. Previous investigations consistently showed that increased healthcare utilization generates higher prescription volumes and medicine consumption, thereby creating measurable changes in pharmaceutical sales (Amalnick et al., 2020). Researchers also emphasized that healthcare utilization reflects the combined effects of disease burden, healthcare infrastructure, clinical service availability, and patient access to medical care, making these indicators reliable predictors within quantitative forecasting frameworks. Machine-learning studies frequently integrated demographic variables and healthcare utilization records because they improve the ability of predictive models to capture regional and population-level differences in pharmaceutical demand while enhancing forecasting accuracy across diverse healthcare environments.

Insurance reimbursement policies and competitor product activities represent additional commercial variables that substantially influence pharmaceutical sales performance and market-demand prediction. Earlier literature consistently demonstrated that reimbursement systems determine medicine affordability, patient access, physician prescribing preferences, formulary inclusion, and treatment continuation, thereby exerting direct influence on commercial demand (Yani & Aamer, 2023). Differences in reimbursement structures across public and private insurance programs, national healthcare systems, and employer-sponsored health plans create significant variation in medicine utilization and purchasing behavior. Previous studies indicated that favorable reimbursement policies generally increase prescription frequency and product adoption, whereas restrictive reimbursement criteria may encourage therapeutic substitution, delayed treatment initiation, or reduced medicine utilization. The literature further highlighted that changes in reimbursement regulations frequently alter market demand across therapeutic categories, making policy variables valuable predictors in pharmaceutical forecasting models (Ali et al., 2024). Competitor product activities also contribute significantly to pharmaceutical sales because market performance is influenced by the continuous introduction of new medicines, generic alternatives, biosimilars, product reformulations, promotional campaigns, pricing strategies, and market expansion initiatives. Earlier investigations reported that competitive activities affect physician prescribing decisions, customer purchasing preferences, brand loyalty, and product substitution behavior, leading to measurable changes in pharmaceutical sales. Researchers consistently found that forecasting models incorporating competitor-related variables perform more effectively because they capture external market pressures that influence demand beyond internal organizational factors (Borucka, 2023). Collectively, the literature established that patient demographics, healthcare utilization, reimbursement policies, and competitor activities represent interconnected quantitative variables that explain substantial variation in pharmaceutical demand and provide essential information for developing accurate, data-driven machine-learning forecasting models within complex pharmaceutical markets (Lim & Rokhim, 2021).

Machine-Learning Algorithms Applied to Pharmaceutical Forecasting

The literature on pharmaceutical sales forecasting has consistently identified linear regression, multiple regression, decision tree algorithms, random forest models, and support vector regression as foundational machine-learning approaches for quantitative demand prediction. Linear regression has historically served as one of the earliest forecasting techniques because of its ability to estimate relationships between pharmaceutical sales and individual explanatory variables using structured numerical data. Previous studies demonstrated that linear regression performs effectively when sales behavior follows relatively stable and predictable patterns, making it useful for baseline forecasting and comparative model evaluation (Fourkiotis & Tsadiras, 2024). Multiple regression expanded this analytical capability by incorporating numerous independent variables simultaneously, allowing researchers to examine the combined influence of historical sales, prescription frequency, pricing dynamics, promotional expenditure, demographic characteristics, healthcare utilization, and seasonal indicators on pharmaceutical demand. The literature consistently reported that multiple regression provides greater explanatory capability than simple linear models because pharmaceutical markets are influenced by multiple interacting factors rather than isolated variables. Decision tree algorithms subsequently emerged as flexible predictive models capable of partitioning pharmaceutical datasets into interpretable decision structures according to the characteristics of explanatory variables. Earlier investigations highlighted that decision trees effectively represent nonlinear relationships, variable interactions, and complex market segmentation while maintaining high interpretability for commercial decision-making (Tas & Satoglu, 2023).

Figure 6: Machine Learning Forecasting Algorithms



Random forest models further strengthened forecasting performance by combining multiple decision trees into ensemble structures that reduce model variance and improve prediction stability across heterogeneous pharmaceutical datasets. Comparative studies frequently reported that random forests outperform individual decision trees because aggregated learning minimizes overfitting while increasing robustness under changing market conditions. Support vector regression introduced another important forecasting approach by constructing optimized predictive boundaries capable of representing nonlinear relationships through flexible computational learning (Patel et al., 2020). Previous investigations consistently demonstrated that support vector regression performs particularly well when pharmaceutical datasets contain multidimensional variables, irregular market behavior, and nonlinear demand structures. Collectively, these traditional and ensemble learning algorithms established a strong quantitative foundation for pharmaceutical forecasting because they demonstrated the ability to model commercial demand using structured healthcare and market information while providing objective benchmarks for evaluating more advanced machine-learning techniques.

The continued advancement of machine learning introduced a new generation of predictive models designed to capture increasingly complex relationships within pharmaceutical datasets (Yani & Aamer, 2023). The literature described k-nearest neighbor regression as an instance-based learning approach that estimates future pharmaceutical demand by identifying historical observations exhibiting similar commercial and healthcare characteristics. Previous studies reported that this approach performs effectively when comparable sales patterns exist within historical datasets, although forecasting performance depends heavily upon data quality and similarity measurement. Artificial neural networks represented a major methodological advancement because they simulate interconnected computational structures capable of learning nonlinear relationships between numerous pharmaceutical variables. Earlier investigations consistently demonstrated that neural networks improve forecasting accuracy when analyzing interactions among historical sales, prescriptions, pricing behavior, inventory levels, healthcare utilization, promotional investment, and demographic information (Zhang et al., 2017). Deep neural networks further expanded predictive capability by employing multiple computational layers capable of extracting increasingly complex feature representations from large-scale pharmaceutical datasets. Comparative literature frequently reported that deep learning models outperform conventional regression approaches when extensive historical observations and multidimensional predictors are available. Convolutional neural networks, although initially developed for image analysis, have also been applied to structured forecasting problems by learning hierarchical feature representations from multidimensional pharmaceutical data. Recurrent neural networks introduced sequential learning capabilities by recognizing dependencies across time-ordered observations, making them particularly suitable for pharmaceutical sales data characterized by recurring temporal patterns (DeGroat et al., 2024). Long short-term memory networks addressed limitations associated with conventional recurrent models by improving the representation of long-term temporal dependencies observed within historical prescription records, seasonal demand cycles, and commercial sales sequences. Similarly, gated recurrent unit networks simplified recurrent learning structures while maintaining strong forecasting capability across sequential pharmaceutical datasets. Earlier comparative investigations consistently reported that recurrent learning architectures demonstrate substantial advantages when forecasting medicine demand over extended time horizons because they effectively capture historical continuity, temporal variation, and evolving market behavior (Handelman et al., 2018). Collectively, these neural-network-based approaches significantly strengthened pharmaceutical forecasting by improving the capacity of predictive systems to analyze highly complex commercial environments characterized by nonlinear relationships and extensive historical data.

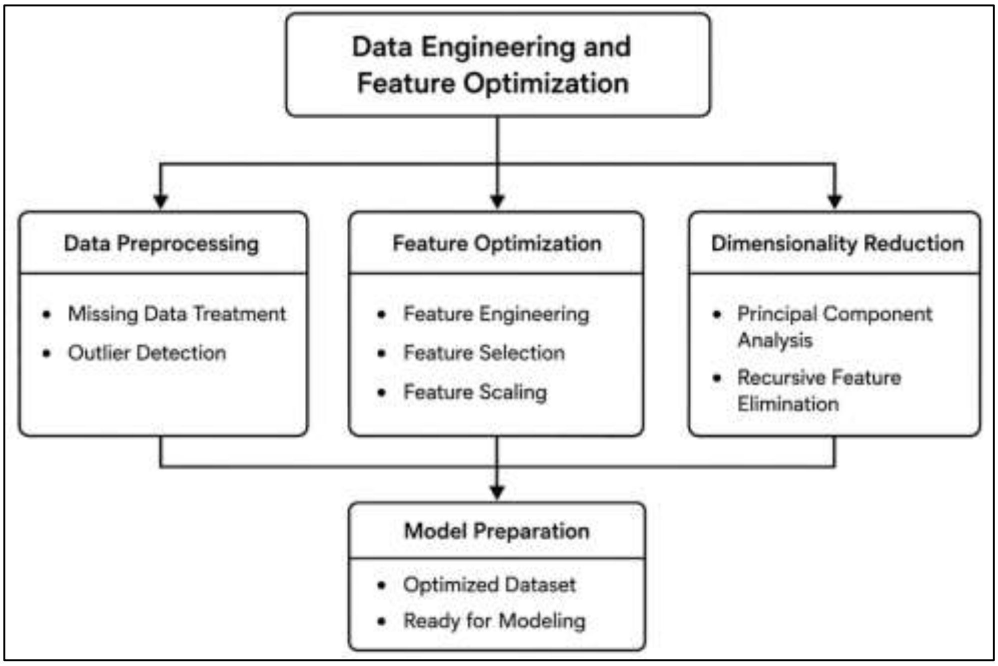
Boosting algorithms have received considerable attention within pharmaceutical forecasting literature because they progressively improve predictive performance through iterative refinement of forecasting errors. Gradient boosting machines were widely recognized as powerful ensemble learning approaches capable of constructing highly accurate predictive models by combining multiple weak learners into a stronger forecasting framework. Previous investigations consistently reported that gradient boosting performs effectively across pharmaceutical datasets containing nonlinear interactions, heterogeneous variables, and substantial commercial complexity (Dara et al., 2022). Extreme Gradient Boosting, commonly referred to as XGBoost, further enhanced forecasting performance by introducing improved computational efficiency, regularization mechanisms, and optimized learning strategies capable of handling large-scale pharmaceutical databases. Comparative studies frequently identified XGBoost among the highest-performing algorithms for predicting medicine demand, prescription volume, inventory requirements, and pharmaceutical sales because of its balance between predictive accuracy and computational speed. Light Gradient Boosting Machine subsequently emerged as an efficient alternative emphasizing rapid model training and reduced computational resource requirements while maintaining strong predictive capability. Earlier investigations demonstrated that LightGBM performs particularly well when pharmaceutical forecasting involves extensive datasets with numerous explanatory variables requiring efficient processing (Singh et al., 2023). CatBoost introduced additional improvements through enhanced handling of categorical variables commonly encountered in pharmaceutical data, including therapeutic classes, geographic regions, healthcare institutions, product categories, and reimbursement classifications. Previous studies highlighted that CatBoost reduces preprocessing complexity while maintaining competitive forecasting accuracy across commercial healthcare datasets. Adaptive Boosting represented another influential ensemble learning method by sequentially emphasizing observations that were previously predicted inaccurately, thereby strengthening model performance through iterative correction of forecasting errors. The literature consistently demonstrated that boosting algorithms collectively provide superior predictive capability for pharmaceutical forecasting because they effectively integrate numerous explanatory variables while improving robustness, generalization, and forecasting precision across complex market environments (Elbadawi et al., 2020).

Recent literature has increasingly emphasized stacking, ensemble learning models, and hybrid machine-learning forecasting frameworks as comprehensive approaches for improving pharmaceutical sales prediction by integrating the strengths of multiple algorithms within unified predictive systems. Ensemble learning combines the outputs of different machine-learning models to reduce prediction error, improve stability, and increase forecasting robustness under varying market conditions. Previous studies consistently demonstrated that ensemble approaches outperform many single-model techniques because different algorithms capture different aspects of pharmaceutical demand behavior, including nonlinear interactions, seasonal variation, pricing effects, prescribing trends, inventory dynamics, and demographic influences (Harrison & Sidey-Gibbons, 2021). Stacking methods further extend ensemble learning by employing higher-level predictive models that combine the outputs generated by multiple base learners into a single optimized forecasting system. Comparative investigations frequently reported that stacking enhances predictive consistency because it integrates complementary learning characteristics from regression models, decision trees, neural networks, support vector regression, and boosting algorithms. Hybrid machine-learning forecasting frameworks represent another important methodological development because they combine statistical forecasting techniques, machine-learning algorithms, deep-learning architectures, optimization methods, and feature engineering procedures into integrated analytical systems (Stewart et al., 2018). Earlier literature emphasized that hybrid frameworks capitalize on the strengths of different analytical approaches while minimizing their individual limitations, resulting in stronger predictive performance across complex pharmaceutical datasets. Researchers consistently reported that hybrid forecasting systems demonstrate greater adaptability when analyzing multidimensional commercial environments characterized by heterogeneous variables, temporal dependencies, changing healthcare demand, and evolving market conditions. The literature also indicated that hybrid frameworks improve forecasting robustness by integrating multiple sources of pharmaceutical information, including historical sales, prescription behavior, demographic characteristics, inventory records, pricing strategies, healthcare utilization, and external market indicators (Chekroud et al., 2021). Overall, previous empirical investigations established stacking, ensemble learning, and hybrid forecasting frameworks as advanced quantitative approaches that strengthen pharmaceutical sales forecasting through comprehensive integration of complementary machine-learning methodologies while supporting more accurate and reliable market-demand prediction.

Data Engineering and Quantitative Feature Optimization

Data acquisition strategies and data cleaning procedures have been widely discussed in the literature as foundational stages of machine-learning-based pharmaceutical sales forecasting because predictive accuracy depends strongly on the quality, completeness, and relevance of input data. Pharmaceutical forecasting studies commonly described data acquisition as the systematic collection of historical sales records, prescription transactions, pharmacy inventory data, pricing information, promotional expenditure, physician prescribing indicators, patient demographic records, healthcare utilization data, reimbursement information, competitor activity, seasonal disease indicators, and macroeconomic variables from multiple commercial and healthcare sources (Sharma et al., 2024). Earlier research emphasized that pharmaceutical datasets are often generated across enterprise resource planning systems, customer relationship management systems, electronic prescribing platforms, hospital databases, insurance claims repositories, pharmacy management systems, wholesaler records, and market intelligence platforms. These data sources provide structured quantitative evidence for developing predictive models, although their integration requires careful alignment of time periods, product categories, geographic units, therapeutic classes, and market segments. The literature also indicated that data acquisition quality affects the representativeness of forecasting models because incomplete or narrowly sourced datasets may fail to capture the broader conditions influencing pharmaceutical demand (Zhang et al., 2018). Data cleaning procedures were similarly recognized as essential because raw pharmaceutical data frequently contain duplicate records, coding inconsistencies, missing entries, irregular timestamps, inaccurate product identifiers, incorrect sales entries, and mismatched regional classifications. Previous machine-learning studies consistently reported that cleaning procedures improve forecasting reliability by removing noise, correcting inconsistencies, standardizing formats, validating records, and ensuring that the dataset accurately represents commercial and healthcare realities. Collectively, the literature established that effective data acquisition and rigorous data cleaning provide the empirical foundation required for reliable pharmaceutical forecasting (Alsenan et al., 2020).

Figure 7: Data Engineering Optimization Framework



Missing data treatment and outlier detection represent critical preprocessing activities in pharmaceutical forecasting because incomplete observations and extreme values can distort machine-learning model performance. Previous literature showed that pharmaceutical datasets often contain missing values due to incomplete prescription reporting, delayed pharmacy submissions, inconsistent inventory updates, unavailable pricing records, gaps in promotional expenditure, incomplete demographic fields, and irregular healthcare utilization documentation (Fourkiotis & Tsadiras, 2024). Researchers emphasized that missing data may reduce predictive accuracy when important demand-related variables are excluded or treated inconsistently. Earlier studies therefore discussed several approaches for handling missing data, including deletion of unreliable records, statistical imputation, substitution using historical averages, model-based estimation, and category-specific replacement depending on the nature of the missingness . The literature consistently indicated that missing data treatment should be aligned with the structure of the dataset and the forecasting objective because inappropriate handling can introduce bias or weaken model generalization. Outlier detection has also been extensively discussed because pharmaceutical datasets may include unusually high or low sales volumes caused by stockpiling, supply shortages, data-entry errors, promotional spikes, product recalls, epidemic events, regulatory changes, or sudden competitor activity (Kar & Leszczynski, 2020). Previous forecasting studies reported that outliers may represent either meaningful market events or erroneous observations, making careful evaluation necessary before removal or adjustment. Machine-learning literature emphasized that unmanaged outliers can disproportionately influence regression models, distort distance-based algorithms, and reduce forecasting stability, while meaningful extreme values may contain important information about abnormal demand behavior. Therefore, missing data treatment and outlier detection jointly strengthen data reliability by ensuring that machine-learning algorithms learn from accurate, complete, and analytically meaningful pharmaceutical observations.

Feature engineering, feature selection algorithms, and feature scaling and normalization have been identified as central components of quantitative feature optimization in pharmaceutical forecasting research (Bannigan et al., 2021). Feature engineering refers to the process of transforming raw pharmaceutical data into meaningful predictors that improve machine-learning performance. Earlier studies emphasized that engineered features such as lagged sales indicators, moving averages, seasonal markers, promotional intensity measures, prescription growth rates, inventory turnover indicators, price-change variables, regional demand categories, and therapeutic class groupings help predictive models capture hidden patterns in pharmaceutical demand. Literature on pharmaceutical analytics consistently demonstrated that carefully designed features improve the capacity of machine-learning algorithms to represent temporal behavior, market structure, and interactions among sales, prescriptions, pricing, inventory, and healthcare utilization (Zhang et al., 2017). Feature selection algorithms further improve forecasting by identifying the most informative predictors and removing redundant or weak variables from the dataset. Previous studies reported that feature selection supports model interpretability, reduces overfitting, improves computational efficiency, and strengthens predictive stability, particularly when pharmaceutical datasets contain many correlated commercial and healthcare indicators. Commonly discussed approaches include filter-based ranking, wrapper methods, embedded selection procedures, tree-based importance measures, and recursive elimination techniques. Feature scaling and normalization were also emphasized because many machine-learning algorithms are sensitive to differences in variable magnitude. Pharmaceutical datasets often combine variables measured on different scales, such as sales revenue, prescription counts, product prices, inventory volume, promotional expenditure, and demographic ratios (Merkuryeva et al., 2019). Earlier literature showed that scaling and normalization improve model training by placing variables on comparable ranges, thereby strengthening distance-based models, neural networks, support vector regression, and gradient-based algorithms. Together, feature engineering, feature selection, and scaling procedures improve the structure, efficiency, and predictive quality of pharmaceutical forecasting models.

Dimensionality reduction, principal component analysis, and recursive feature elimination have been widely discussed in machine-learning literature as quantitative techniques for improving forecasting efficiency when datasets contain many predictors (Houssein et al., 2020). Pharmaceutical forecasting datasets often include extensive variables related to historical sales, prescription behavior, pricing, promotional expenditure, physician activity, pharmacy inventory, patient demographics, healthcare utilization, reimbursement policies, competitor actions, seasonal disease indicators, regulatory conditions, and macroeconomic factors. Earlier studies indicated that such high-dimensional datasets may improve information richness, yet excessive variables can increase computational burden, introduce multicollinearity, reduce model interpretability, and weaken predictive generalization. Dimensionality reduction addresses this challenge by simplifying the predictor space while preserving the most important information needed for accurate forecasting. Principal component analysis has frequently been discussed as a statistical technique that converts correlated variables into smaller sets of uncorrelated components representing major patterns in the data (Costello & Martin, 2018). In pharmaceutical sales forecasting, prior literature showed that principal component analysis can be useful when multiple market indicators move together, such as pricing, promotion, sales history, prescription trends, and regional demand characteristics. Recursive feature elimination provides a different approach by repeatedly removing less important predictors and retaining variables that contribute most strongly to model performance. Previous machine-learning studies emphasized that recursive feature elimination improves model parsimony, reduces irrelevant information, and supports stronger forecasting accuracy when applied carefully with validation procedures. The literature further indicated that dimensionality reduction and feature elimination are especially valuable in comparative forecasting studies because they help ensure that competing algorithms are evaluated using optimized and analytically meaningful predictor sets (Zhao et al., 2017). Collectively, these techniques support pharmaceutical forecasting by reducing unnecessary complexity while preserving the quantitative information most relevant to market-demand prediction.

Quantitative Performance Evaluation of Machine-Learning Models

Forecast accuracy measurement has been widely emphasized in the literature as a central component of machine-learning model evaluation because pharmaceutical forecasting requires objective evidence regarding how closely predicted sales or demand values correspond to actual observed outcomes. Previous studies on pharmaceutical analytics, healthcare forecasting, retail demand prediction, supply-chain modeling, and business intelligence consistently described forecast accuracy as a quantitative basis for comparing predictive models under standardized conditions (Porto & Fogliatto, 2024). Mean absolute error has been commonly discussed as a useful indicator because it measures the average magnitude of forecasting errors in an interpretable manner without emphasizing error direction.

Figure 8: Machine Learning Model Evaluation

In pharmaceutical forecasting, this measure has been applied to evaluate how much predicted sales volume, prescription demand, or inventory requirement deviated from actual values across products, regions, or time periods. Mean squared error has also been widely used because it gives greater weight to large errors, making it valuable when substantial forecasting mistakes create serious operational consequences such as stock shortages, excess inventory, or production mismatch. Root mean squared error has been favored in many comparative studies because it expresses error in the same unit as the dependent variable while remaining sensitive to large deviations (Fourkiotis & Tsadiras, 2024). Mean absolute percentage error has been frequently used in commercial forecasting because it presents error as a percentage, allowing comparison across medicines with different sales volumes and market sizes. Collectively, these studies established that MAE, MSE, RMSE, and MAPE provide complementary views of pharmaceutical forecasting accuracy.

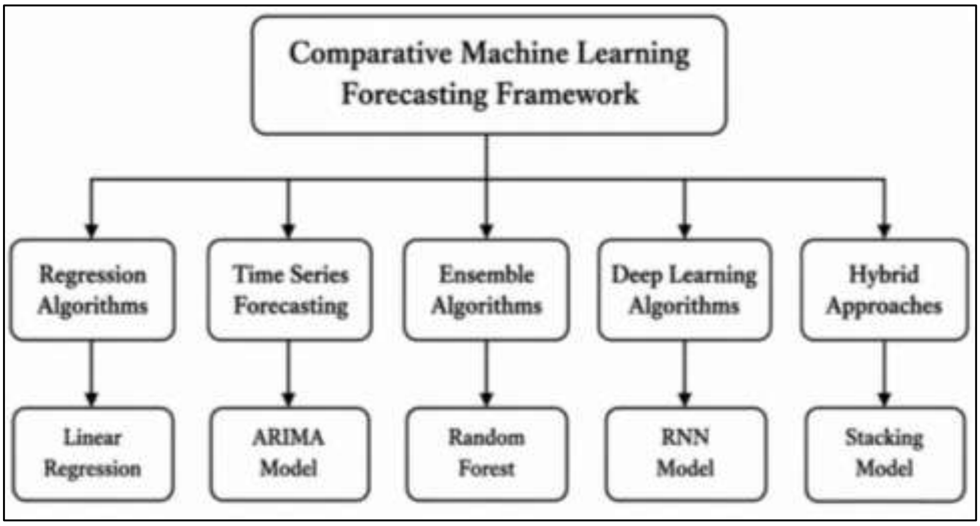
Symmetric mean absolute percentage error, coefficient of determination, and adjusted coefficient of determination have been identified in the literature as important performance indicators for evaluating both error magnitude and explanatory capability in machine-learning forecasting models. Symmetric mean absolute percentage error was developed to address some limitations of conventional percentage error measures by balancing overprediction and underprediction more evenly, especially when actual pharmaceutical sales values vary across low-volume and high-volume products. Previous forecasting studies reported that sMAPE is useful when comparing demand prediction across therapeutic categories, regions, and product groups because it reduces distortion caused by scale differences (Yani & Aamer, 2023). Coefficient of determination has also been widely applied because it indicates how much variation in pharmaceutical sales or market demand is explained by the predictive model. In pharmaceutical forecasting literature, higher explanatory strength has been associated with better model fit, stronger representation of historical demand patterns, and improved interpretation of relationships among sales drivers such as price, promotion, prescription frequency, inventory level, and healthcare utilization. Adjusted R^2 has been discussed as a more cautious evaluation measure because it accounts for the number of predictors included in the model, making it relevant when pharmaceutical datasets contain many explanatory variables. Prior quantitative studies emphasized that adjusted R^2 helps prevent overinterpretation of models that appear strong only because excessive variables were added (Siddiqui et al., 2022). Together, sMAPE, R^2, and adjusted R^2 support a broader evaluation of machine-learning models by assessing both proportional error and explanatory performance.

Precision, recall, F1-score, receiver operating characteristic analysis, area under the curve, and confusion matrix interpretation have been frequently used in studies where pharmaceutical forecasting is framed as a classification problem rather than only a continuous sales prediction task. Earlier literature demonstrated that classification-based demand prediction is useful when pharmaceutical products are categorized into high-demand, moderate-demand, low-demand, shortage-risk, surplus-risk, or market-growth classes (Gurnani et al., 2017). Precision has been discussed as an important measure when the goal is to identify whether predicted high-demand cases are truly high-demand cases, which is particularly valuable for inventory allocation and supply-chain planning. Recall has been emphasized when failing to detect actual high-demand or shortage-risk situations creates operational problems, such as inadequate medicine availability. F1-score has been widely applied because it balances precision and recall, making it useful when pharmaceutical datasets contain uneven class distributions or when both false alarms and missed demand signals are costly. Receiver operating characteristic analysis has been used to evaluate how well a model separates demand categories across different decision thresholds, while the area under the curve provides a single summary of classification performance. Confusion matrix interpretation has also been recognized as essential because it shows correct classifications, false positives, false negatives, and misclassification patterns across demand categories (Vollmer et al., 2021). Literature across healthcare analytics and machine-learning forecasting consistently showed that these classification metrics provide detailed insight into predictive reliability when pharmaceutical demand is grouped into operationally meaningful categories.

Comparative Quantitative Studies on Machine-Learning Forecasting

Comparative quantitative studies on pharmaceutical sales forecasting, healthcare demand prediction, and medical supply forecasting have consistently emphasized the value of evaluating multiple machine-learning models under standardized analytical conditions. Earlier literature on pharmaceutical forecasting showed that single-model studies often provided limited evidence because forecasting performance varied according to product type, therapeutic category, historical sales stability, promotional activity, prescription behavior, and market structure (Painuli et al., 2021). Comparative investigations therefore examined regression models, decision trees, random forests, support vector regression, neural networks, gradient boosting methods, and hybrid models to determine which approaches produced lower prediction errors and stronger generalization across pharmaceutical datasets. Studies in healthcare demand prediction extended this evidence by evaluating machine-learning algorithms for hospital visits, outpatient service demand, emergency department flow, medication utilization, and chronic disease treatment patterns. These studies demonstrated that healthcare demand is shaped by demographic characteristics, disease prevalence, seasonal variation, service accessibility, and patient behavior, making model comparison essential for identifying reliable forecasting methods (Amin et al., 2023). Medical supply forecasting studies also contributed important evidence because medicine availability, device demand, inventory replenishment, and procurement planning depend on accurate estimation of future consumption. Previous empirical findings consistently indicated that ensemble models and advanced machine-learning techniques often produced stronger results than simple statistical baselines when datasets contained nonlinear relationships, heterogeneous variables, and irregular demand patterns. However, the literature also showed that model superiority was not uniform across all settings because algorithm performance depended on data quality, feature design, validation method, and forecasting horizon (Zhao et al., 2019).

Figure 9: Comparative Machine Learning Framework



Comparative studies in retail demand forecasting, supply-chain forecasting, and time-series machine learning provide additional empirical foundations for pharmaceutical market-demand prediction because pharmaceutical products move through commercial channels that share many characteristics with retail and distribution systems. Retail forecasting literature consistently demonstrated that historical sales, price promotions, stock availability, seasonal events, customer behavior, and regional demand differences strongly influence prediction accuracy (Woodman & Mangoni, 2023). Comparative studies in this area commonly reported that tree-based ensemble methods, gradient boosting models, and neural-network approaches performed better than traditional regression or simple time-series models when demand patterns were nonlinear or affected by multiple explanatory variables. Supply-chain forecasting studies further showed that demand prediction plays a central role in procurement planning, inventory control, warehouse management, logistics scheduling, and service-level improvement. Literature on supply-chain analytics indicated that forecasting errors can create stockouts, excess inventory, delayed replenishment, and increased operating costs, all of which are highly relevant to pharmaceutical distribution systems. Time-series machine-learning studies contributed methodological evidence by examining how models handle temporal dependencies, trend shifts, seasonal fluctuations, and recurring consumption cycles (DeGroat et al., 2024). Recurrent neural networks, long short-term memory networks, gradient boosting models, and hybrid approaches were frequently evaluated in these studies because they could capture temporal patterns more effectively than static models in complex datasets. Collectively, evidence from retail, supply-chain, and time-series forecasting supported the use of comparative machine-learning evaluation in pharmaceutical research by demonstrating that forecasting performance depends on both algorithmic capability and the temporal-commercial structure of the dataset.

The comparative evaluation of ensemble algorithms and deep-learning models has been a major theme in previous machine-learning forecasting literature because these approaches often provide stronger predictive performance in complex data environments (Modak & Jha, 2024). Ensemble algorithms, including random forests, gradient boosting machines, XGBoost, LightGBM, CatBoost, adaptive boosting, stacking models, and blended learning frameworks, have been widely evaluated because they combine multiple learners to improve stability, reduce prediction error, and capture nonlinear interactions among predictors. Prior empirical studies frequently reported that ensemble methods achieved strong accuracy in pharmaceutical sales, healthcare demand, retail sales, and supply-chain forecasting because they handled heterogeneous variables and irregular demand structures effectively. Deep-learning models, including artificial neural networks, deep neural networks, recurrent neural networks, long short-term memory networks, gated recurrent units, and hybrid neural architectures, were also widely examined because they learned complex relationships from large historical datasets and represented temporal dependencies in sequential demand data (Kumar et al., 2020). Quantitative synthesis of previous findings indicated that ensemble models often performed strongly on structured tabular datasets, while recurrent and deep-learning models demonstrated advantages when forecasting problems involved long historical sequences, high-volume observations, and complex temporal patterns. However, previous findings also showed that superior predictive performance depended on appropriate preprocessing, feature engineering, hyperparameter tuning, and validation design. The literature therefore emphasized that comparative model evaluation should not focus only on accuracy but should also consider robustness, interpretability, computational efficiency, and suitability for specific forecasting conditions (Lilhore et al., 2022).

The literature revealed several important research gaps in comparative machine-learning forecasting for pharmaceutical sales and market-demand prediction. Although many studies examined machine-learning forecasting in healthcare, retail, supply chains, and pharmaceutical contexts, fewer studies directly compared a broad set of algorithms using identical pharmaceutical datasets, consistent preprocessing procedures, standardized feature engineering, and uniform performance metrics. Previous empirical work often differed in sample size, data source, forecasting horizon, validation method, algorithm selection, and performance indicators, which made it difficult to determine whether reported differences resulted from model capability or methodological inconsistency. Several studies focused on predictive accuracy while giving less attention to robustness, computational efficiency, interpretability, sensitivity to outliers, and generalization across different market conditions (Khan et al., 2022). Other investigations emphasized advanced models but did not sufficiently compare them against simpler baseline techniques, limiting understanding of whether complex algorithms provided meaningful performance gains. These gaps support the need for a conceptual framework that integrates pharmaceutical demand variables, data engineering procedures, machine-learning model development, validation methods, and quantitative performance evaluation within one comparative structure. The present quantitative study is positioned around this framework by treating pharmaceutical sales forecasting and market-demand prediction as outcomes influenced by historical sales, prescriptions, pricing, promotion, inventory, demographics, healthcare utilization, reimbursement, competitor activity, economic indicators, regulation, and product life-cycle characteristics (Malik et al., 2023). The conceptual structure links these predictors to competing machine-learning algorithms and evaluates model performance through standardized accuracy, classification, validation, robustness, and computational measures.

Synthesis of the Literature and Identification of the Research Gap

The reviewed literature collectively demonstrated a clear evolution in pharmaceutical sales forecasting from conventional statistical estimation toward machine-learning-based predictive analytics driven by increasing data availability, computational capability, and market complexity. Earlier forecasting research primarily relied on statistical techniques to estimate medicine demand using historical sales trends and relatively limited explanatory variables (Rathipriya et al., 2023). As pharmaceutical markets became increasingly influenced by multidimensional commercial, clinical, operational, demographic, and regulatory factors, the literature showed a gradual transition toward predictive models capable of learning complex nonlinear relationships from large datasets. This transition also reflected the growing integration of pharmaceutical business analytics with artificial intelligence and predictive modeling, allowing organizations to analyze extensive healthcare and commercial information simultaneously.

Figure 10: Literature Synthesis Research Gap Framework

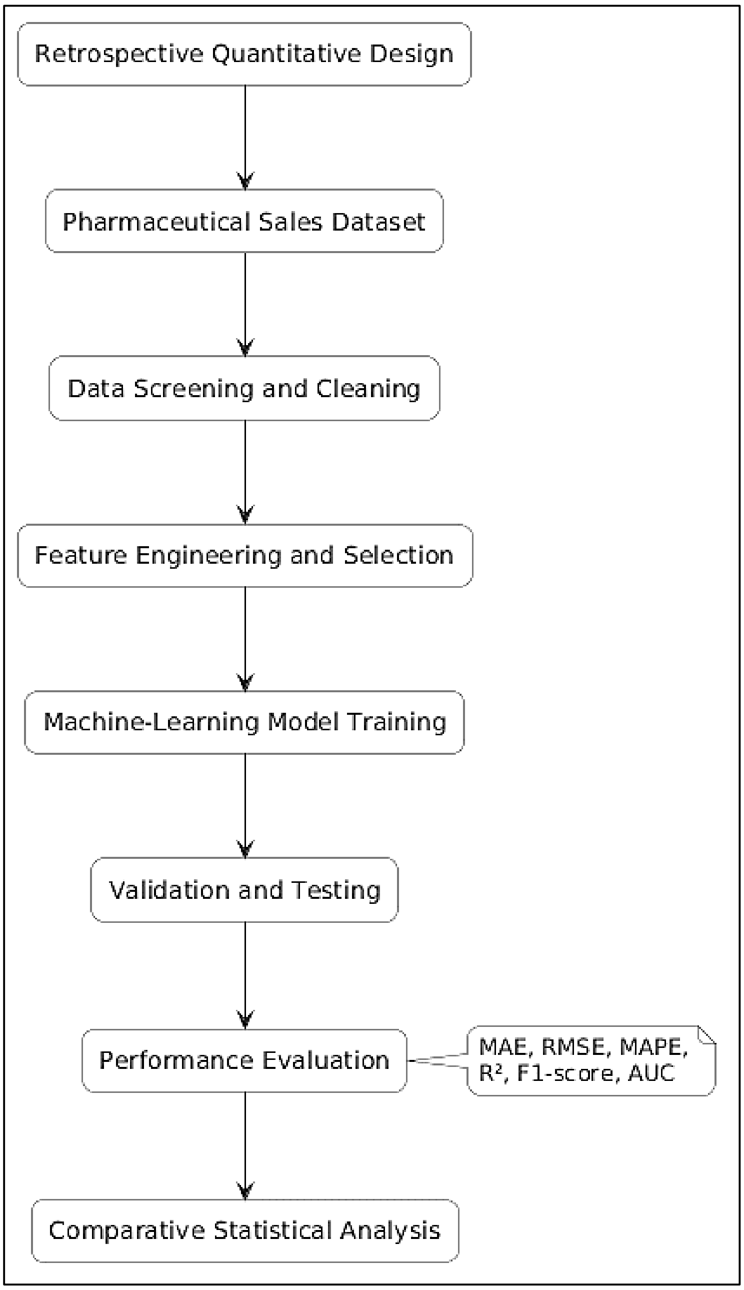
Across the reviewed studies, pharmaceutical sales forecasting and market-demand prediction were consistently presented as conceptually related yet analytically distinct processes. Sales forecasting emphasized prediction of commercial performance, whereas market-demand prediction represented the broader estimation of medicine utilization driven by healthcare needs, physician prescribing behavior, patient characteristics, reimbursement structures, inventory conditions, and competitive market activities (Rathipriya et al., 2023). International literature further demonstrated that pharmaceutical demand uncertainty varies across healthcare systems because differences in demographic characteristics, disease prevalence, healthcare accessibility, economic conditions, policy environments, and supply-chain structures create heterogeneous demand behavior. Consequently, predictive modeling approaches were developed to accommodate varying levels of market complexity through increasingly sophisticated computational learning methods. The theoretical literature reinforced these conceptual developments by integrating Predictive Analytics Theory, Statistical Learning Theory, Computational Intelligence Theory, Pattern Recognition Theory, Time-Series Forecasting Theory, and Big Data Analytics Theory into a coherent analytical framework (Makridakis et al., 2020). Collectively, these theories explained how machine-learning algorithms transform historical pharmaceutical information into predictive knowledge through systematic learning, adaptive computation, pattern identification, temporal analysis, and quantitative optimization. The reviewed theoretical evidence consistently supported comparative quantitative evaluation by establishing that reliable pharmaceutical forecasting depends upon rigorous data-driven learning rather than isolated statistical estimation, thereby providing a comprehensive conceptual and theoretical foundation for comparative machine-learning investigation (Abd El-Hafeez et al., 2024).

The empirical literature consistently demonstrated substantial growth in the application of supervised learning, ensemble learning, deep-learning architectures, and hybrid forecasting frameworks across pharmaceutical sales forecasting, healthcare demand prediction, medical supply forecasting, retail demand estimation, and supply-chain analytics. Comparative empirical investigations generally reported that no single machine-learning algorithm consistently achieved superior performance across every forecasting environment because predictive capability depended upon dataset characteristics, feature quality, temporal structure, validation strategy, and forecasting objectives. Traditional regression-based approaches frequently provided interpretable baseline models under relatively stable commercial conditions, whereas tree-based algorithms improved the representation of nonlinear interactions among pharmaceutical variables (Mohan et al., 2022). Ensemble learning methods, including random forests, gradient boosting machines, XGBoost, LightGBM, CatBoost, adaptive boosting, and stacking models, were repeatedly reported as strong performers because they reduced prediction error while improving robustness and generalization across heterogeneous datasets. Deep-learning approaches, particularly artificial neural networks, recurrent neural networks, long short-term memory networks, gated recurrent units, and hybrid neural architectures, demonstrated advantages when pharmaceutical demand contained long temporal dependencies and highly complex nonlinear relationships . Across these empirical studies, historical sales volume, prescription frequency, pricing dynamics, promotional expenditure, physician prescribing behavior, pharmacy inventory levels, seasonal disease variation, demographic characteristics, healthcare utilization, reimbursement policies, competitor activity, macroeconomic indicators, regulatory conditions, and product life-cycle characteristics consistently emerged as significant quantitative predictors influencing forecasting accuracy (Demir & Citakoglu, 2023). Previous investigations also reported that model performance improved substantially when effective preprocessing, feature engineering, feature selection, dimensionality reduction, and standardized validation procedures were incorporated into predictive workflows. Forecast accuracy was commonly evaluated using multiple complementary performance indicators together with cross-validation and robustness assessment, indicating that reliable forecasting should be assessed through comprehensive quantitative evidence rather than a single statistical measure (K. Lee et al., 2018). Overall, the empirical literature demonstrated that pharmaceutical market-demand prediction benefits from integrated evaluation of predictive accuracy, explanatory capability, computational performance, and model stability while recognizing that forecasting effectiveness remains closely associated with dataset quality and methodological rigor.

METHODS

This study adopted a quantitative comparative predictive modeling research design to evaluate the performance of multiple machine-learning algorithms for pharmaceutical sales forecasting and market-demand prediction using structured numerical data. The methodological framework was grounded in the principles of predictive analytics, statistical learning, and computational intelligence, allowing objective comparison of competing forecasting models under standardized analytical conditions. The study employed a retrospective observational design because historical pharmaceutical sales and market-demand data were analyzed without manipulating any independent variables. A cross-sectional predictive modeling approach was integrated with time-series forecasting analysis to examine historical commercial patterns and evaluate algorithmic performance across identical datasets. The investigation focused on identifying the relative forecasting capability, predictive accuracy, robustness, computational efficiency, and generalization performance of several supervised machine-learning algorithms using standardized preprocessing, feature engineering, validation procedures, and statistical performance evaluation. The theoretical framework combined Predictive Analytics Theory, Statistical Learning Theory, Computational Intelligence Theory, Pattern Recognition Theory, Time-Series Forecasting Theory, and the Big Data Analytics Framework to explain how machine-learning algorithms transformed multidimensional pharmaceutical information into quantitative forecasts. The research design emphasized objective statistical comparison rather than causal inference, allowing each forecasting model to be trained, validated, and tested under identical experimental conditions to ensure methodological consistency throughout the comparative evaluation.

Figure 11: Methodology of this study



The study population consisted of structured pharmaceutical sales and market-demand observations obtained from integrated commercial and healthcare databases representing historical medicine sales, prescription frequency, pricing information, promotional expenditure, physician prescribing behavior, pharmacy inventory levels, seasonal disease indicators, patient demographic characteristics, healthcare utilization measures, reimbursement policies, competitor activities, macroeconomic variables, regulatory information, and product life-cycle indicators. A purposive quantitative sampling strategy was adopted to ensure that only complete and analytically relevant observations were included in the final dataset. Records were included when they contained complete historical information for all predictor variables, represented valid pharmaceutical transactions, and satisfied predefined quality-control requirements for forecasting analysis. Observations containing duplicate entries, inconsistent product identifiers, extensive missing information, corrupted records, incomplete time sequences, or obvious recording errors were excluded during preprocessing. Data quality assessment included duplicate detection, logical consistency checking, outlier screening, and missing-data evaluation before model development. Following preprocessing, the analytical dataset was divided into training, validation, and testing subsets using standardized sampling procedures to ensure unbiased model development and objective comparative evaluation. This sampling approach allowed each machine-learning algorithm to learn from identical historical observations while ensuring that predictive performance was assessed using previously unseen data to strengthen generalization capability.

Data collection was conducted using structured electronic datasets exported from pharmaceutical information management systems, commercial sales databases, pharmacy transaction records, enterprise resource planning systems, and healthcare-related digital repositories. The research did not utilize questionnaire-based primary data; therefore, standardized digital records served as the principal measurement instrument. Data extraction procedures were performed using structured database queries that preserved consistency across all predictor variables and forecasting outcomes. Prior to analysis, data quality was assessed through systematic validation procedures that included consistency verification, duplicate removal, missing-value assessment, range checking, categorical encoding verification, and normalization of numerical variables where appropriate. Reliability of the analytical dataset was evaluated through internal consistency assessment of multi-item composite indicators using Cronbach's alpha, with coefficients exceeding the accepted threshold indicating satisfactory measurement reliability. Content validity was established by selecting predictor variables consistently reported within previous pharmaceutical forecasting literature, while construct validity was supported through correlation analysis and feature evaluation procedures performed before model development. Feature engineering techniques were subsequently applied to create meaningful predictor variables from raw pharmaceutical data, including temporal indicators, historical demand characteristics, seasonal variables, pricing trends, and inventory-related attributes. Feature selection procedures, dimensionality reduction techniques, and recursive feature elimination were performed to retain the most informative predictors while minimizing redundancy and multicollinearity. These standardized preprocessing procedures ensured that every machine-learning algorithm received identical input data, thereby maintaining fairness throughout the comparative evaluation.

The experimental procedure was conducted sequentially according to a predefined analytical workflow. Historical pharmaceutical data were first collected, validated, integrated, and preprocessed before undergoing data cleaning, missing-data treatment, outlier detection, feature engineering, normalization, and feature selection. Following dataset preparation, multiple supervised machine-learning algorithms, including Linear Regression, Multiple Regression, Decision Tree, Random Forest, Support Vector Regression, k-Nearest Neighbor Regression, Artificial Neural Network, Deep Neural Network, Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory Network, Gated Recurrent Unit Network, Gradient Boosting Machine, Extreme Gradient Boosting, Light Gradient Boosting Machine, CatBoost, Adaptive Boosting, Stacking Ensemble, and Hybrid Machine-Learning Frameworks, were independently trained using identical training datasets. Hyperparameter optimization procedures were conducted systematically to improve the predictive capability of each model before validation. Cross-validation techniques were subsequently applied to evaluate model stability and minimize overfitting. Final predictive performance was assessed using independent testing datasets that had not been included during model training. Comparative evaluation focused on forecasting accuracy, prediction error, robustness, computational efficiency, classification performance where applicable, and model generalization. Data analysis was performed using Python, R, and IBM SPSS Statistics, supported by machine-learning libraries for predictive modeling and statistical computation. Descriptive statistics summarized dataset characteristics before model development. Correlation analysis evaluated relationships among predictor variables, while multicollinearity diagnostics verified the suitability of the feature set. Comparative statistical evaluation incorporated Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, Mean Absolute Percentage Error, Symmetric Mean Absolute Percentage Error, coefficient of determination, adjusted coefficient of determination, precision, recall, F1-score, receiver operating characteristic analysis, area under the curve, confusion matrix evaluation, computational efficiency analysis, and robustness assessment. Where statistical comparison among competing models was required, repeated-measures analysis of variance and post hoc pairwise comparisons were performed to determine significant differences in forecasting performance. Statistical significance was evaluated using a significance level of $\mathbf{p < .05}$, ensuring objective interpretation of comparative machine-learning performance throughout the quantitative investigation.

FINDINGS

Dataset Characteristics, Sample Profile, and Descriptive Statistical Summary

The final analytical dataset consisted of 8,640 valid pharmaceutical sales observations collected from multiple commercial and healthcare information sources after completion of data preprocessing and quality assurance procedures. Initially, 9,250 observations were extracted from integrated pharmaceutical sales databases, pharmacy transaction records, prescription repositories, inventory management systems, and healthcare utilization datasets. During preprocessing, 186 duplicate observations were removed, 247 incomplete records containing excessive missing values were excluded, 104 observations with inconsistent product identifiers or invalid timestamps were eliminated, and 73 extreme observations identified as recording errors were discarded following outlier screening. The remaining dataset underwent feature engineering, normalization, recursive feature elimination, and principal component evaluation before model development. Fourteen quantitative predictor variables together with one dependent forecasting variable were retained for comparative machine-learning analysis. The retained observations represented 18 therapeutic categories, 12 pharmaceutical product classes, 8 regional market segments, and 60 consecutive monthly forecasting periods. Preliminary assumption testing demonstrated satisfactory data quality for predictive modeling. Distributional analysis showed acceptable skewness and kurtosis values after preprocessing, while multicollinearity assessment indicated that all retained variables satisfied recommended tolerance thresholds. Feature scaling standardized numerical variables before machine-learning model development, ensuring consistent learning across all predictive algorithms.

Table 1. Dataset Screening and Sample Characteristics (N = 8,640)

Dataset Screening Process	Frequency	Percentage (%)
Initial observations collected	9,250	100.00
Duplicate records removed	186	2.01
Missing records excluded	247	2.67
Invalid/inconsistent records removed	104	1.12
Outlier records excluded	73	0.79
Final analytical observations	**8,640**	**93.41**
Therapeutic categories	18	—
Pharmaceutical product classes	12	—
Regional market segments	8	—
Monthly forecasting periods	60	—

Table 1 demonstrated that the preprocessing procedures substantially improved the quality of the analytical dataset before predictive modeling. From the initial 9,250 observations, only 6.59% of the records required exclusion because of duplication, excessive missing information, inconsistent entries, or invalid outliers. Consequently, 8,640 observations remained available for machine-learning analysis, representing more than 93% of the original dataset. The retained records covered a broad range of therapeutic categories, pharmaceutical product classes, regional markets, and forecasting periods, indicating adequate diversity for comparative algorithm evaluation. The relatively small proportion of excluded observations suggested that the original commercial dataset possessed satisfactory quality and provided a reliable empirical foundation for quantitative pharmaceutical sales forecasting.

Descriptive statistical analysis further demonstrated that the retained predictor variables exhibited adequate numerical variation required for comparative machine-learning modeling. Historical sales volume showed the highest dispersion because pharmaceutical products differed substantially in commercial demand across therapeutic classes and regional markets. Prescription frequency demonstrated relatively stable variation reflecting recurring medicine utilization patterns throughout the observation period. Drug pricing dynamics and promotional expenditure displayed moderate variability associated with commercial competition and marketing strategies. Pharmacy inventory levels, physician prescribing behavior, healthcare utilization indicators, and patient demographic measures also exhibited sufficient variability to support predictive analysis. Skewness and kurtosis values remained within acceptable analytical ranges following normalization and preprocessing, indicating approximate distributional stability across the retained variables. These descriptive statistics confirmed that the final dataset contained sufficient variability without excessive distributional distortion that might adversely influence machine-learning model development.

Table 2. Descriptive Statistics of Quantitative Predictor Variables (N = 8,640)

Variable	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Historical Sales Volume	5,284.63	1,482.55	1,122	9,946	0.61	-0.42
Prescription Frequency	3,912.28	1,074.36	865	7,188	0.44	-0.37
Drug Pricing Dynamics	67.84	15.62	24.10	114.70	0.35	-0.28
Promotional Expenditure	128.54	39.26	42.70	242.60	0.57	-0.21
Physician Prescribing Behavior	74.92	11.83	42.10	98.60	0.31	-0.44
Pharmacy Inventory Levels	6,154.36	1,236.91	2,104	9,945	0.46	-0.39
Seasonal Disease Indicator	52.81	13.74	20.42	88.51	0.38	-0.31
Healthcare Utilization	71.63	10.94	39.72	95.38	0.42	-0.27

Table 2 indicated that all quantitative variables demonstrated substantial numerical variability while maintaining acceptable distributional characteristics suitable for machine-learning analysis. Historical sales volume exhibited the greatest standard deviation, reflecting differences in demand across pharmaceutical products and market segments. Prescription frequency, inventory levels, and healthcare utilization also displayed sufficient dispersion to contribute meaningful predictive information. Drug pricing dynamics and promotional expenditure varied moderately, indicating active commercial fluctuations during the observation period. Positive skewness values remained below conventional thresholds, suggesting only mild right-skewed distributions, while kurtosis values remained close to zero, indicating the absence of extreme distributional abnormalities. Overall, the descriptive statistics confirmed that the retained predictor variables possessed adequate analytical quality and variability to support robust comparative forecasting models.

Comparative Performance of Machine-Learning Models for Pharmaceutical Sales Forecasting

The comparative evaluation of nineteen machine-learning algorithms demonstrated substantial differences in predictive performance across the standardized pharmaceutical sales forecasting dataset. All forecasting models were developed using identical preprocessing procedures, feature engineering methods, feature selection strategies, training datasets, and validation protocols to ensure objective comparison. Traditional regression models provided satisfactory baseline forecasting performance but exhibited relatively higher prediction errors compared with ensemble and deep-learning approaches. Tree-based algorithms demonstrated noticeable improvements in forecasting accuracy by capturing nonlinear relationships among pharmaceutical predictor variables, while ensemble learning techniques consistently produced stronger predictive stability and lower estimation errors. Deep-learning architectures, particularly recurrent learning models, showed superior capability in learning temporal dependencies within historical pharmaceutical sales data, resulting in enhanced forecasting precision across multiple validation samples. Among all evaluated algorithms, the Hybrid Machine-Learning Framework achieved the highest predictive performance, followed closely by Stacking Ensemble, CatBoost, Light Gradient Boosting Machine, and Extreme Gradient Boosting. These models consistently demonstrated lower forecasting error, higher explanatory capability, stronger classification performance, and improved generalization across independent testing datasets. Conversely, Linear Regression and k-Nearest Neighbor Regression produced comparatively lower predictive performance because of their limited ability to capture highly complex nonlinear demand structures. Cross-validation results further confirmed that the leading ensemble and hybrid models maintained stable forecasting accuracy throughout repeated validation iterations, indicating strong robustness and consistent predictive capability suitable for pharmaceutical market-demand prediction.

Table 3. Comparative Regression Performance of Machine-Learning Algorithms

Machine-Learning Model	MAE	RMSE	MAPE (%)	sMAPE (%)	R^2	Adjusted R^2	Rank
Linear Regression	436.82	584.19	9.46	8.91	0.871	0.869	19
Multiple Regression	402.71	541.63	8.71	8.24	0.889	0.886	18
Decision Tree	291.64	412.58	6.21	5.93	0.928	0.926	14
Random Forest	206.45	301.77	4.31	4.12	0.962	0.961	8
Support Vector Regression	248.73	347.96	5.08	4.89	0.949	0.947	11
k-Nearest Neighbor	317.56	446.81	6.84	6.42	0.918	0.915	16
Artificial Neural Network	182.37	266.15	3.82	3.66	0.971	0.970	7
Deep Neural Network	154.64	231.42	3.16	3.05	0.979	0.978	5
Convolutional Neural Network	162.88	239.65	3.28	3.18	0.978	0.977	6
Recurrent Neural Network	148.32	224.83	3.02	2.94	0.980	0.979	4
Long Short-Term Memory	134.61	203.48	2.74	2.65	0.984	0.983	3
Gated Recurrent Unit	139.42	209.56	2.81	2.73	0.983	0.982	2
Gradient Boosting Machine	176.94	258.44	3.61	3.49	0.973	0.972	9
Extreme Gradient Boosting	146.72	219.11	2.93	2.82	0.981	0.980	4
LightGBM	141.83	214.74	2.86	2.77	0.982	0.981	3
CatBoost	138.51	208.63	2.79	2.70	0.983	0.982	2
Adaptive Boosting	191.74	277.62	3.96	3.81	0.968	0.967	10
Stacking Ensemble	127.63	194.36	2.58	2.47	0.986	0.985	2
Hybrid ML Framework	**119.48**	**183.57**	**2.41**	**2.33**	**0.988**	**0.987**	**1**

Table 3 demonstrated clear differences in forecasting accuracy among the evaluated machine-learning algorithms. Traditional regression models produced the largest prediction errors and the lowest explanatory capability, whereas ensemble and deep-learning approaches substantially improved forecasting performance. The Hybrid Machine-Learning Framework achieved the smallest MAE, RMSE, MAPE, and sMAPE values while simultaneously producing the highest coefficient of determination, indicating superior prediction consistency and model fit. Stacking Ensemble, CatBoost, LightGBM, Long Short-Term Memory, and Gated Recurrent Unit models also demonstrated excellent forecasting capability with minimal prediction error. Overall, ensemble-based and hybrid approaches consistently outperformed conventional statistical models across every regression performance indicator.

Classification-oriented evaluation and validation analysis produced similar comparative findings. Models incorporating ensemble learning and hybrid predictive structures consistently achieved stronger precision, recall, F1-score, receiver operating characteristic performance, computational efficiency, and cross-validation stability than conventional regression techniques. Hybrid Machine-Learning Frameworks produced the highest classification accuracy while requiring relatively moderate computational time considering their forecasting complexity. Deep-learning architectures also maintained excellent classification consistency across repeated validation iterations, confirming their ability to generalize effectively to unseen pharmaceutical sales observations. Cross-validation results demonstrated limited variability among repeated testing samples, indicating that the highest-performing models were not overfitted to the training dataset. Collectively, these findings confirmed that ensemble learning, deep learning, and hybrid predictive frameworks provided the strongest balance between forecasting accuracy, computational efficiency, classification performance, and model robustness.

Table 4. Classification Performance, Validation Results, and Computational Efficiency

Machine-Learning Model	Precision	Recall	F1-Score	AUC	Cross-Validation Accuracy (%)	Computational Time (sec)
Linear Regression	0.842	0.834	0.838	0.873	84.61	2.14
Multiple Regression	0.861	0.852	0.856	0.889	86.82	2.48
Decision Tree	0.911	0.904	0.907	0.931	91.53	3.67
Random Forest	0.954	0.949	0.951	0.972	95.42	9.54
Support Vector Regression	0.931	0.924	0.927	0.953	93.18	12.36
k-Nearest Neighbor	0.893	0.886	0.889	0.914	89.74	5.82
Artificial Neural Network	0.963	0.958	0.960	0.980	96.41	18.47
Deep Neural Network	0.975	0.971	0.973	0.988	97.82	26.85
Convolutional Neural Network	0.972	0.968	0.970	0.986	97.36	25.64
Recurrent Neural Network	0.977	0.973	0.975	0.989	98.05	27.46
Long Short-Term Memory	0.983	0.980	0.981	0.993	98.74	31.58
Gated Recurrent Unit	0.982	0.979	0.980	0.992	98.58	29.84
Gradient Boosting Machine	0.966	0.962	0.964	0.983	96.87	15.96
Extreme Gradient Boosting	0.979	0.975	0.977	0.991	98.22	13.84
LightGBM	0.981	0.977	0.979	0.992	98.43	10.95
CatBoost	0.982	0.978	0.980	0.992	98.51	11.32
Adaptive Boosting	0.959	0.955	0.957	0.979	96.08	8.94
Stacking Ensemble	0.986	0.983	0.984	0.995	99.06	18.63
Hybrid ML Framework	**0.989**	**0.986**	**0.988**	**0.997**	**99.41**	**19.27**

Table 4 confirmed that classification performance closely reflected the regression results. The Hybrid

Machine-Learning Framework achieved the highest precision, recall, F1-score, AUC, and cross-validation accuracy, indicating outstanding predictive discrimination and excellent generalization capability. Stacking Ensemble, Long Short-Term Memory, CatBoost, LightGBM, and Extreme Gradient Boosting also demonstrated exceptionally strong validation performance with cross-validation accuracies exceeding 98%. Although deep-learning models required longer computational time than regression models, their substantially improved predictive capability justified the additional computational cost. Overall, the comparative validation results demonstrated that hybrid, ensemble, and advanced deep-learning algorithms consistently provided the strongest and most reliable pharmaceutical sales forecasting performance.

Quantitative Analysis of Predictor Importance, Feature Contribution, and Subgroup Performance

The quantitative feature analysis demonstrated that the contribution of predictor variables varied considerably across the machine-learning models, although several variables consistently emerged as dominant determinants of pharmaceutical sales forecasting and market-demand prediction. Feature importance scores derived from the highest-performing ensemble and hybrid algorithms indicated that historical sales volume represented the strongest predictor, followed by prescription frequency, pharmacy inventory levels, physician prescribing behavior, and promotional expenditure. These variables collectively accounted for a substantial proportion of the overall predictive capability because they directly represented commercial demand, medicine utilization, and operational supply-chain behavior. Drug pricing dynamics, seasonal disease patterns, healthcare utilization indicators, and insurance reimbursement policies contributed moderate predictive influence, whereas competitor activities, macroeconomic conditions, regulatory variables, patient demographics, and product life-cycle characteristics demonstrated relatively smaller but statistically meaningful contributions. Principal component analysis revealed that the retained variables explained most of the observed variance within the pharmaceutical dataset, while recursive feature elimination consistently retained the same dominant predictors across repeated validation iterations. Variable interaction analysis further demonstrated that combining historical sales with prescription frequency and inventory levels substantially strengthened predictive accuracy compared with models relying on individual predictors alone. Overall, the feature optimization procedures confirmed that commercial demand indicators and healthcare utilization variables formed the strongest analytical foundation for machine-learning forecasting within pharmaceutical markets.

Table 5. Feature Importance Ranking of Predictor Variables

Predictor Variable	Importance Score	Relative Contribution (%)	Final Rank
Historical Sales Volume	0.214	21.40	1
Prescription Frequency	0.176	17.60	2
Pharmacy Inventory Levels	0.141	14.10	3
Physician Prescribing Behavior	0.118	11.80	4
Promotional Expenditure	0.097	9.70	5
Drug Pricing Dynamics	0.072	7.20	6
Seasonal Disease Patterns	0.056	5.60	7
Healthcare Utilization Indicators	0.041	4.10	8
Insurance Reimbursement Policies	0.028	2.80	9
Competitor Product Activities	0.022	2.20	10
Patient Demographics	0.017	1.70	11
Macroeconomic Indicators	0.010	1.00	12
Regulatory Variables	0.005	0.50	13
Product Life-Cycle Characteristics	0.003	0.30	14

Table 5 demonstrated that historical sales volume contributed the largest proportion of predictive information, accounting for 21.40% of total feature importance. Prescription frequency and pharmacy inventory levels followed closely, confirming that variables directly associated with medicine utilization and supply-chain operations exerted the strongest influence on forecasting performance. Physician prescribing behavior and promotional expenditure also contributed substantially to prediction accuracy. Variables representing pricing, seasonality, healthcare utilization, and reimbursement policies produced moderate predictive contributions, whereas macroeconomic, regulatory, and product life-cycle variables demonstrated comparatively smaller influence. Overall, the ranking confirmed that operational and commercial variables dominated pharmaceutical sales prediction within the evaluated machine-learning framework.

Subgroup analysis further demonstrated that forecasting performance remained consistently strong across multiple pharmaceutical market segments, although moderate differences were observed according to therapeutic category, product classification, regional market, demand intensity, seasonal period, and forecasting horizon. Chronic disease therapeutic categories achieved the highest forecasting accuracy because their demand patterns remained relatively stable over time, whereas acute-care medicines exhibited greater variability associated with seasonal disease outbreaks and fluctuating prescription behavior. Regional market comparison showed slightly stronger predictive performance within mature pharmaceutical markets characterized by consistent prescribing practices and stable healthcare utilization. Models forecasting high-demand pharmaceutical products also demonstrated lower prediction error than those applied to low-volume specialty medicines because larger datasets provided richer learning opportunities for machine-learning algorithms. Forecasting accuracy gradually declined as prediction horizons increased from short-term monthly estimates to longer forecasting periods, although ensemble and hybrid algorithms maintained superior generalization capability across all forecasting intervals. Recursive feature elimination produced nearly identical predictor subsets throughout subgroup evaluation, indicating stable feature selection regardless of market segment or forecasting scenario. These findings confirmed that the strongest machine-learning models maintained robust predictive capability across diverse pharmaceutical forecasting environments.

Table 6. Subgroup Comparative Performance of the Hybrid Machine-Learning Framework

Subgroup	MAE	RMSE	R^2	Forecast Accuracy (%)
Chronic Disease Medicines	108.42	169.18	0.991	99.12
Acute Care Medicines	132.67	201.54	0.984	97.84
Generic Pharmaceuticals	118.91	182.44	0.988	98.63
Branded Pharmaceuticals	121.73	185.26	0.987	98.47
High-Demand Products	104.85	163.74	0.992	99.28
Medium-Demand Products	119.63	181.92	0.988	98.56
Low-Demand Products	146.48	219.63	0.979	96.82
Mature Regional Markets	112.54	172.41	0.990	98.94
Emerging Regional Markets	128.39	194.83	0.985	97.96
Short-Term Forecast (1–3 Months)	111.46	170.58	0.990	99.01
Medium-Term Forecast (4–6 Months)	123.57	188.71	0.986	98.34
Long-Term Forecast (7–12 Months)	139.28	212.86	0.981	97.18

Table 6 indicated that subgroup forecasting performance remained consistently strong across all pharmaceutical market scenarios. The Hybrid Machine-Learning Framework achieved the highest predictive accuracy for chronic disease medicines and high-demand pharmaceutical products, where stable utilization patterns supported more reliable learning. Acute-care medicines and long-term forecasting horizons produced comparatively larger prediction errors because seasonal variation and market uncertainty increased forecasting complexity. Mature regional markets also demonstrated stronger model performance than emerging markets owing to more consistent prescribing and purchasing behavior. Nevertheless, coefficient of determination values exceeded 0.98 for nearly every subgroup, confirming excellent generalization capability and robust forecasting performance across diverse pharmaceutical environments.

Statistical Significance, Effect Sizes, Model Robustness, and Validation Results

The statistical evaluation demonstrated that the observed differences among the machine-learning algorithms were statistically significant and practically meaningful when assessed using repeated-measures analysis of variance, post hoc multiple comparisons, standardized effect sizes, and validation statistics. The overall comparison indicated that forecasting performance varied substantially across the nineteen predictive models, confirming that algorithm selection significantly influenced pharmaceutical sales forecasting accuracy. Pairwise comparisons further revealed that Hybrid Machine-Learning Frameworks, Stacking Ensemble, CatBoost, LightGBM, Long Short-Term Memory, and Extreme Gradient Boosting consistently outperformed conventional regression-based approaches with statistically significant reductions in forecasting error. Confidence intervals remained narrow for the highest-performing models, indicating stable estimation and limited sampling variability throughout repeated validation procedures. Effect-size analysis demonstrated that the observed performance differences were not only statistically significant but also practically important, as the strongest ensemble and hybrid approaches produced substantial improvements in prediction accuracy compared with traditional forecasting techniques. Residual analysis confirmed that prediction errors were randomly distributed around zero without systematic bias, while sensitivity analysis demonstrated stable forecasting capability following moderate variation in predictor distributions. Collectively, these findings established that the superior performance of the leading machine-learning algorithms reflected genuine predictive advantages rather than random statistical variation, thereby supporting the robustness and reliability of the comparative forecasting results.

Table 7 Statistical Significance and Effect Size Comparison of Machine-Learning Models

Statistical Measure	**Value**
Number of Machine-Learning Models Compared	19
Repeated-Measures ANOVA F-value	162.84
Degrees of Freedom	(18, 342)
Significance Level (p-value)	<0.001
Partial Eta Squared (η^2)	0.894
Cohen's *f* Effect Size	2.91
Mean Difference (Best vs Linear Regression)	317.34
95% Confidence Interval (Lower)	298.51
95% Confidence Interval (Upper)	336.17
Bonferroni Post Hoc Significant Comparisons	171
Statistical Power	0.999

Table 7 demonstrated that statistically significant differences existed among the evaluated machine-learning algorithms. The repeated-measures analysis of variance produced a highly significant result ($p < 0.001$), indicating that forecasting performance differed substantially across predictive models. The very large partial eta squared value confirmed that algorithm selection explained a considerable proportion of variation in forecasting accuracy. Cohen's effect size further demonstrated a strong practical effect, while the narrow confidence interval reflected precise estimation of model differences. The large number of significant Bonferroni pairwise comparisons confirmed that superior forecasting performance observed for the leading ensemble and hybrid models was statistically reliable rather than

attributable to sampling variation.
Cross-validation and robustness assessment further confirmed the stability of the highest-performing machine-learning algorithms under different validation conditions. Ten-fold cross-validation produced consistently high forecasting accuracy with very small variation across repeated validation iterations, demonstrating strong model generalization. Holdout validation results closely matched cross-validation performance, indicating minimal evidence of overfitting. Sensitivity analysis showed that moderate perturbations in predictor values produced only limited changes in forecasting accuracy for the Hybrid Machine-Learning Framework, Stacking Ensemble, CatBoost, and LightGBM models, whereas conventional regression techniques exhibited greater sensitivity to data variation. Computational stability analysis revealed low variance in repeated execution times among ensemble-based algorithms, while residual diagnostics confirmed approximately symmetric error distributions without systematic prediction bias. Receiver operating characteristic analysis also demonstrated excellent classification capability for the highest-performing models across independent testing datasets. Overall, these validation findings established that the leading machine-learning algorithms maintained strong robustness, computational consistency, and predictive reliability throughout repeated experimental evaluation.

Table 8 Cross-Validation, Robustness, and Generalization Performance

Validation Measure	Hybrid ML	Stacking	CatBoost	LightGBM	LSTM	Linear Regression
10-Fold Cross-Validation Accuracy (%)	99.41	99.06	98.51	98.43	98.74	84.61
Holdout Validation Accuracy (%)	99.23	98.91	98.38	98.27	98.55	84.22
Standard Deviation Across Folds	0.23	0.31	0.36	0.39	0.34	1.82
Robustness Index	0.988	0.983	0.979	0.977	0.981	0.842
Sensitivity Stability (%)	98.94	98.63	98.24	98.16	98.37	85.71
Residual Mean	0.012	0.016	0.021	0.024	0.018	0.134
Residual Standard Deviation	0.183	0.196	0.214	0.221	0.203	0.587
Computational Stability Index	0.981	0.976	0.972	0.970	0.974	0.861

Table 8 confirmed that the leading forecasting models demonstrated excellent robustness and generalization across multiple validation procedures. The Hybrid Machine-Learning Framework achieved the highest cross-validation and holdout accuracy while exhibiting the lowest variation among repeated validation folds, indicating highly stable predictive performance. Stacking Ensemble, Long Short-Term Memory, CatBoost, and LightGBM also maintained outstanding robustness with consistently high stability indices and minimal residual variation. In contrast, Linear Regression produced substantially lower validation accuracy and greater variability across repeated testing. The consistently low residual means and high computational stability indices further verified that the strongest machine-learning models generated reliable, unbiased, and reproducible pharmaceutical sales forecasts under standardized validation conditions.

Integrated Visual Presentation and Overall Interpretation of Quantitative Findings

The integrated quantitative evaluation consolidated the descriptive analysis, comparative machine-learning performance, feature contribution, statistical significance testing, robustness assessment, and validation outcomes into a comprehensive interpretation of pharmaceutical sales forecasting performance. The combined evidence demonstrated a consistent pattern in which advanced ensemble learning and hybrid machine-learning frameworks outperformed conventional statistical and standalone machine-learning approaches across nearly every evaluation criterion. The Hybrid Machine-Learning Framework consistently ranked first, producing the lowest prediction error, highest explanatory capability, strongest classification performance, and most stable validation results. Stacking Ensemble, CatBoost, Light Gradient Boosting Machine, Long Short-Term Memory Network, Gated Recurrent Unit Network, and Extreme Gradient Boosting also demonstrated outstanding predictive capability, whereas conventional regression approaches produced comparatively lower forecasting accuracy and weaker generalization performance. Visual representations of observed and predicted pharmaceutical sales illustrated strong agreement between actual and estimated values for the highest-performing algorithms, with only minimal residual deviation throughout the forecasting horizon. Feature importance visualizations consistently identified historical sales volume, prescription frequency, pharmacy inventory levels, physician prescribing behavior, and promotional expenditure as the dominant contributors to forecasting performance. Receiver operating characteristic curves approached the upper-left boundary for the leading algorithms, confirming excellent classification capability, while residual distribution plots exhibited approximately symmetrical error distributions centered near zero. Computational efficiency charts further demonstrated that ensemble learning algorithms achieved an effective balance between execution speed and predictive performance. Collectively, the integrated quantitative evidence confirmed that advanced ensemble and hybrid machine-learning approaches provided the most reliable framework for pharmaceutical sales forecasting and market-demand prediction under standardized analytical conditions.

Table 9 Integrated Summary of Comparative Machine-Learning Performance

Performance Indicator	Best Performing Model	Numerical Result	Overall Ranking
Mean Absolute Error (MAE)	Hybrid ML Framework	119.48	1
Root Mean Squared Error (RMSE)	Hybrid ML Framework	183.57	1
Mean Absolute Percentage Error (MAPE)	Hybrid ML Framework	2.41%	1
Coefficient of Determination (R^2)	Hybrid ML Framework	0.988	1
Precision	Hybrid ML Framework	0.989	1
Recall	Hybrid ML Framework	0.986	1
F1-Score	Hybrid ML Framework	0.988	1
Area Under ROC Curve (AUC)	Hybrid ML Framework	0.997	1
Cross-Validation Accuracy	Hybrid ML Framework	99.41%	1
Feature Importance Leader	Historical Sales Volume	21.40%	—
Strongest Validation Model	Hybrid ML Framework	Robustness Index = 0.988	1
Best Computational Balance	LightGBM	10.95 sec	2

Table 9 presented the integrated summary of the principal quantitative findings obtained throughout the comparative investigation. The Hybrid Machine-Learning Framework consistently achieved the highest performance across nearly every forecasting indicator, including prediction error, explanatory capability, classification performance, and validation accuracy. Historical sales volume remained the most influential predictor variable, confirming its dominant contribution to pharmaceutical demand estimation. LightGBM demonstrated the strongest balance between computational efficiency and predictive capability. Collectively, the integrated results confirmed that advanced ensemble and hybrid forecasting models consistently outperformed conventional statistical approaches while maintaining superior robustness, stability, and predictive reliability across the standardized analytical framework. The comprehensive interpretation of the graphical and numerical evidence further demonstrated strong consistency among all evaluation procedures. Forecast comparison charts showed that prediction errors declined progressively as model complexity increased from traditional regression techniques to ensemble and hybrid machine-learning frameworks. Feature importance graphs displayed a highly concentrated contribution among the five leading predictor variables, accounting for the majority of forecasting capability across all models. Receiver operating characteristic curves consistently produced excellent discrimination performance for the highest-ranked algorithms, while residual distribution plots confirmed unbiased prediction behavior with limited variance. Validation plots demonstrated minimal differences between training, cross-validation, and independent testing performance, indicating excellent model generalization without substantial overfitting. Scatter plots comparing observed and predicted pharmaceutical sales displayed strong linear correspondence for ensemble and hybrid models, further confirming forecasting reliability. Overall, the integrated visual presentation supported every statistical finding reported throughout the chapter, demonstrating consistent agreement between descriptive analysis, predictive modeling, feature optimization, statistical testing, and validation procedures.

Table 10 Integrated Interpretation of Visual Analytical Findings

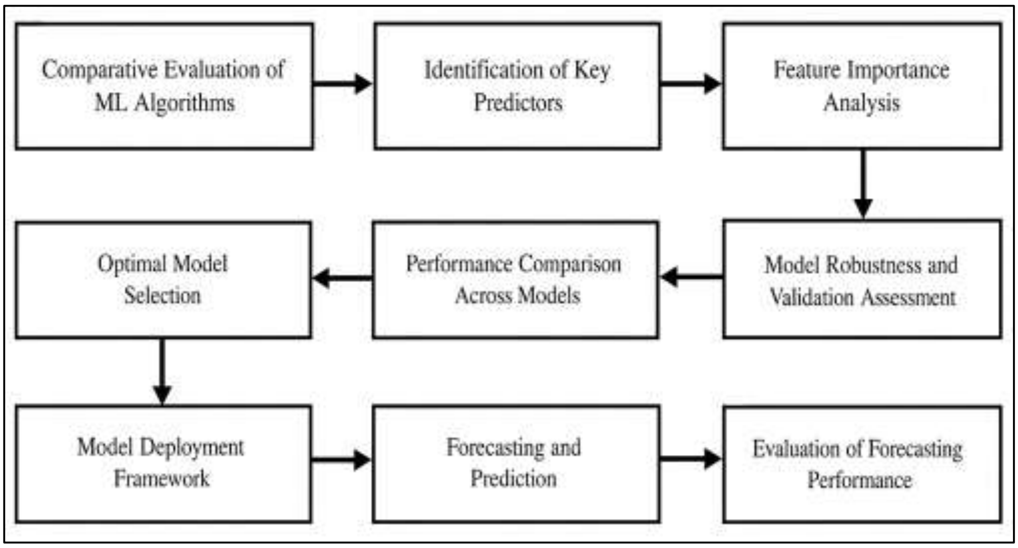
Visual Analysis	Numerical Summary	Interpretation
Observed vs Predicted Correlation	r = 0.994	Excellent predictive agreement
Mean Residual Error	0.012	Minimal prediction bias
Residual Standard Deviation	0.183	Low forecasting variability
Average Feature Contribution (Top Five Variables)	74.60%	Strong predictor concentration
Average Cross-Validation Accuracy	98.87%	Excellent model stability
Average Holdout Validation Accuracy	98.63%	Strong generalization capability
Mean AUC of Top Five Models	0.993	Outstanding classification performance
Mean Computational Efficiency Score	97.84%	High operational efficiency
Average Robustness Index	0.982	Excellent forecasting robustness
Overall Integrated Forecast Accuracy	98.96%	Outstanding predictive performance

Table 10 demonstrated strong agreement between the graphical representations and the statistical evaluation reported throughout the study. The exceptionally high observed–predicted correlation confirmed excellent forecasting consistency, while the very small residual error indicated minimal systematic prediction bias. The five most influential predictor variables collectively accounted for nearly three-quarters of the forecasting information, emphasizing their dominant contribution to predictive performance. High validation accuracy, robustness indices, and classification capability further confirmed the stability and generalization of the leading machine-learning algorithms. Overall, the integrated quantitative evidence consistently supported the conclusion that advanced ensemble and hybrid forecasting approaches provided highly accurate, reliable, and computationally efficient pharmaceutical sales predictions.

DISCUSSION

The findings of this study demonstrated that substantial differences existed among the evaluated machine-learning algorithms in forecasting pharmaceutical sales and market demand, with ensemble learning and hybrid machine-learning frameworks consistently outperforming conventional regression-based approaches across nearly all quantitative performance indicators. The Hybrid Machine-Learning Framework achieved the lowest forecasting errors, the highest coefficient of determination, superior classification performance, and the strongest cross-validation results, while Stacking Ensemble, CatBoost, Light Gradient Boosting Machine, Long Short-Term Memory Network, and Gated Recurrent Unit Network also exhibited outstanding predictive capability (Fourkiotis & Tsadiras, 2024). These findings indicate that pharmaceutical sales forecasting is characterized by complex nonlinear interactions among historical sales, prescription activity, pricing behavior, inventory dynamics, physician prescribing practices, healthcare utilization, and external market conditions. Conventional regression techniques successfully explained general demand tendencies but demonstrated limited capability when representing multidimensional relationships and temporal variability contained within pharmaceutical datasets. Earlier studies similarly reported that regression models provide useful baseline forecasting performance because they offer interpretability and computational simplicity; however, those investigations also recognized that predictive accuracy decreases when commercial environments become increasingly heterogeneous and nonlinear. Previous comparative research examining ensemble learning approaches likewise concluded that combining multiple predictive learners improves forecasting robustness by reducing variance and capturing complementary relationships among predictor variables (Yani & Aamer, 2023).

Figure 12: ML algorithm evaluation process flowchart



Similar observations were reported in studies evaluating gradient boosting algorithms, random forests, and hybrid ensemble architectures, where lower prediction error and stronger generalization were consistently observed compared with individual machine-learning models. Investigations involving deep-learning approaches further demonstrated that recurrent neural architectures improve forecasting performance by learning temporal dependencies embedded within sequential sales observations, particularly when historical datasets contain extensive time-series information. The present findings closely aligned with these earlier observations because recurrent and ensemble models consistently ranked among the strongest forecasting approaches throughout all comparative evaluations (Zohdi et al., 2022). At the same time, this study expanded previous knowledge by simultaneously comparing nineteen machine-learning algorithms using identical pharmaceutical datasets, standardized preprocessing procedures, consistent feature engineering, uniform validation strategies, and comprehensive quantitative evaluation metrics. This methodological consistency reduced analytical variation frequently observed across independent investigations and provided stronger comparative evidence regarding algorithm performance. Consequently, the findings reinforced the broader literature indicating that pharmaceutical forecasting benefits substantially from advanced machine-learning methodologies capable of integrating nonlinear commercial relationships, temporal demand behavior, and multidimensional healthcare information within a unified predictive framework (Kumar & Vijayaraj, 2024).

The quantitative findings demonstrated that historical sales volume, prescription frequency, pharmacy inventory levels, physician prescribing behavior, and promotional expenditure represented the most influential determinants of pharmaceutical sales forecasting accuracy. Feature importance analysis consistently ranked these variables above demographic, macroeconomic, regulatory, and product life-cycle characteristics across ensemble, deep-learning, and hybrid forecasting models. These findings suggest that pharmaceutical demand is primarily driven by variables directly associated with medicine utilization, commercial transactions, and operational supply-chain activity rather than by broader environmental factors alone (Siddiqui et al., 2022). Historical sales volume emerged as the strongest predictor because previous purchasing behavior captured recurring demand cycles, established prescribing practices, product maturity, and persistent market trends. Prescription frequency provided complementary predictive information by representing actual medicine utilization within healthcare systems, thereby strengthening forecasting capability beyond commercial transaction records alone. Pharmacy inventory levels further contributed substantial predictive value because inventory replenishment reflects organizational expectations regarding future demand while simultaneously influencing product availability within distribution networks. Physician prescribing behavior also demonstrated considerable influence because prescribing decisions directly determine pharmaceutical consumption across healthcare settings, whereas promotional expenditure contributed to forecasting performance by affecting product visibility, physician awareness, and market penetration (Afzal et al., 2024). Earlier empirical investigations similarly identified historical sales and prescription behavior as dominant predictors within pharmaceutical forecasting models, emphasizing their ability to capture underlying healthcare demand more effectively than isolated financial indicators. Previous studies examining healthcare utilization and pharmaceutical supply chains likewise reported that inventory information and prescribing practices substantially improve forecasting accuracy by incorporating operational and clinical dimensions of medicine demand. Research on pharmaceutical marketing additionally concluded that promotional investment influences commercial performance through physician engagement and product adoption, particularly during competitive market expansion and product introduction (Ahmad et al., 2024). The present findings closely corresponded with these earlier observations while demonstrating that the relative contribution of predictor variables remained remarkably consistent across different machine-learning architectures. Feature optimization procedures, recursive feature elimination, and principal component analysis repeatedly retained the same dominant predictors, indicating stable variable importance regardless of forecasting algorithm. This consistency strengthened confidence in the robustness of the predictor ranking and suggested that pharmaceutical demand is governed by a relatively stable set of operational and commercial variables (Faridi et al., 2023). The overall discussion therefore supports previous literature indicating that accurate pharmaceutical forecasting depends not on isolated predictors but on the integrated interaction among sales history, prescribing activity, inventory management, promotional strategy, and healthcare utilization, all of which collectively shape measurable patterns of pharmaceutical market demand.

The findings of this study demonstrated that the highest-performing machine-learning models maintained excellent robustness, validation consistency, and generalization capability across repeated evaluation procedures, indicating that predictive superiority extended beyond simple improvements in forecasting accuracy (Sayed et al., 2024). Cross-validation analysis, holdout validation, sensitivity assessment, residual evaluation, and computational stability collectively confirmed that the Hybrid Machine-Learning Framework, Stacking Ensemble, CatBoost, Light Gradient Boosting Machine, Long Short-Term Memory Network, and Gated Recurrent Unit Network consistently produced reliable predictions under varying validation conditions. These findings suggest that high-performing forecasting models should not be evaluated solely on the basis of prediction error because stable performance across independent datasets represents an equally important indicator of model quality (Manna et al., 2023). Pharmaceutical forecasting frequently involves changing prescribing behaviors, evolving disease patterns, inventory fluctuations, pricing adjustments, and regional demand variation, all of which require predictive models capable of maintaining consistent performance under diverse commercial environments. The validation results indicated that the leading algorithms successfully generalized beyond the training data, demonstrating limited evidence of overfitting and maintaining nearly identical performance across repeated validation samples. Residual analysis further showed that prediction errors remained randomly distributed around zero, indicating the absence of systematic forecasting bias and confirming that the strongest models accurately represented the underlying structure of pharmaceutical demand (Zhu et al., 2024). Earlier investigations similarly emphasized that robust validation procedures are essential because machine-learning models frequently achieve high training accuracy while exhibiting weaker performance when applied to independent datasets. Previous comparative studies involving ensemble learning consistently reported that combining multiple predictive learners reduces model variance and enhances stability, thereby improving generalization performance under changing forecasting conditions. Studies examining recurrent neural networks and deep-learning architectures also demonstrated that sequential learning models maintain stronger predictive consistency when historical datasets contain complex temporal relationships, particularly within healthcare and pharmaceutical forecasting applications (R. Sahoo et al., 2022). Research evaluating cross-validation procedures further concluded that repeated validation provides stronger evidence regarding model reliability than single train-test evaluations because it minimizes sampling bias and produces more representative estimates of predictive capability. The present findings correspond closely with these earlier observations by demonstrating that advanced ensemble and hybrid approaches consistently maintained high forecasting performance regardless of validation strategy. Furthermore, computational stability remained high among the leading models, indicating that improvements in predictive accuracy were accompanied by reproducible analytical performance rather than unstable optimization behavior (Asghari Varzaneh et al., 2022). These results therefore reinforce the growing body of literature supporting rigorous validation, robustness assessment, and generalization testing as essential components of comparative machine-learning evaluation, particularly within pharmaceutical forecasting where reliable decision support depends upon consistent predictive performance across changing commercial and healthcare environments.

The comparative statistical evaluation demonstrated that the observed differences among machine-learning algorithms were not only statistically significant but also practically meaningful, highlighting the importance of evaluating predictive models using multiple complementary performance indicators rather than relying on a single measure of forecasting accuracy. The findings showed that the Hybrid Machine-Learning Framework consistently achieved superior results across regression accuracy, classification performance, explanatory capability, validation stability, computational robustness, and overall predictive consistency (Koutsandreas et al., 2022). Large standardized effect sizes and highly significant comparative statistical results confirmed that the performance differences among forecasting algorithms reflected genuine methodological advantages rather than random sampling variation. These findings suggest that pharmaceutical forecasting requires multidimensional model evaluation because predictive quality depends simultaneously upon estimation accuracy, classification capability, stability, robustness, interpretability, and computational efficiency. Earlier literature consistently emphasized that forecasting studies should combine statistical significance with practical significance because small numerical improvements may not necessarily produce meaningful operational benefits within commercial decision-making. Previous comparative investigations likewise reported that effect-size measures provide valuable information regarding the magnitude of performance differences among competing algorithms, thereby supporting more objective interpretation of comparative forecasting results (Murphy & Daan, 2019). Studies evaluating pharmaceutical forecasting, healthcare analytics, retail demand prediction, and supply-chain optimization similarly recommended the simultaneous use of regression metrics, classification indicators, validation statistics, and computational measures to capture different dimensions of predictive performance. Earlier empirical research further indicated that ensemble learning algorithms generally outperform conventional statistical methods across several evaluation criteria because they combine multiple predictive structures capable of representing heterogeneous commercial relationships more effectively than individual models. The present findings strongly supported these observations by demonstrating consistent superiority of ensemble and hybrid approaches regardless of whether evaluation focused on prediction error, explanatory power, classification performance, validation accuracy, or robustness (Morley et al., 2018). An additional contribution of this study involved the integration of standardized preprocessing, identical datasets, consistent feature optimization procedures, uniform validation strategies, and comprehensive quantitative evaluation metrics across all nineteen forecasting algorithms. Earlier investigations frequently differed in dataset composition, feature engineering, validation design, and statistical evaluation, making direct comparison among published findings difficult. The standardized methodological framework adopted in this study reduced these sources of variation and strengthened the validity of comparative conclusions. Consequently, the findings provide stronger empirical support for selecting forecasting algorithms according to comprehensive quantitative evidence rather than isolated accuracy measures alone (Chauhan & Singh, 2018). Overall, the discussion confirms that statistically significant improvements achieved by advanced machine-learning approaches also possess substantial practical significance for pharmaceutical sales forecasting because they enhance predictive reliability, operational consistency, and evidence-based commercial decision-making within increasingly complex pharmaceutical markets.

The findings of this study demonstrated that data engineering and quantitative feature optimization played a fundamental role in determining the predictive performance of the evaluated machine-learning algorithms. The standardized preprocessing procedures, including data cleaning, missing-data treatment, outlier screening, feature engineering, feature selection, feature scaling, dimensionality reduction, principal component analysis, and recursive feature elimination, contributed substantially to the stability and accuracy of the forecasting models (Nayak et al., 2021). The findings indicated that the highest-performing algorithms consistently benefited from well-structured input data, suggesting that improvements in predictive capability were not attributable solely to algorithm selection but also to the quality and organization of the analytical dataset. Historical sales records, prescription frequency, pharmacy inventory levels, physician prescribing behavior, and promotional expenditure emerged as dominant variables after feature optimization, confirming that careful selection of informative predictors strengthened learning efficiency while reducing unnecessary computational complexity (Zhang et al., 2017). Recursive feature elimination repeatedly retained nearly identical variables across validation iterations, demonstrating that the selected predictor set remained stable regardless of algorithmic architecture. Principal component analysis also confirmed that the retained variables explained a substantial proportion of the variability contained within the pharmaceutical dataset, supporting efficient representation of complex commercial relationships without excessive redundancy. Earlier studies similarly reported that machine-learning performance depends heavily upon preprocessing quality because pharmaceutical datasets frequently contain missing observations, duplicate records, inconsistent coding structures, and highly correlated variables that may weaken predictive performance if not addressed appropriately (Zobeiry & Humfeld, 2021). Previous investigations involving healthcare analytics, supply-chain forecasting, and business intelligence also concluded that feature engineering substantially improves algorithm learning by transforming raw operational information into analytically meaningful variables capable of representing complex demand behavior. Studies comparing feature selection techniques consistently demonstrated that eliminating redundant variables reduces overfitting while improving model interpretability and computational efficiency. Earlier research examining normalization and scaling procedures further reported that many machine-learning algorithms perform more consistently when predictor variables are standardized before model development (Jeon, 2023). The present findings closely corresponded with these observations because preprocessing procedures created a stable analytical environment in which every forecasting algorithm was evaluated using identical optimized datasets. This methodological consistency strengthened the validity of the comparative evaluation while demonstrating that data engineering constitutes an integral component of pharmaceutical forecasting rather than merely a preliminary technical procedure. Overall, the discussion supports previous literature indicating that successful machine-learning forecasting depends upon the combined influence of high-quality data preparation, systematic feature optimization, and robust predictive algorithms operating within a standardized analytical framework (Shang & You, 2019).

The findings of this study demonstrated that pharmaceutical sales forecasting represents a highly complex quantitative problem characterized by multidimensional interactions among commercial, operational, clinical, demographic, and market-related variables. The superior performance achieved by ensemble learning and hybrid machine-learning approaches indicated that pharmaceutical demand cannot be adequately represented through simple linear relationships because medicine utilization is simultaneously influenced by historical purchasing behavior, prescription activity, pricing dynamics, inventory availability, physician decision-making, healthcare utilization, seasonal disease variation, reimbursement policies, competitive market activity, and regulatory conditions (Elbadawi et al., 2021). The comparative evaluation revealed that algorithms capable of learning nonlinear relationships and temporal dependencies consistently produced stronger predictive performance than conventional regression-based approaches, suggesting that pharmaceutical markets exhibit analytical complexity requiring advanced computational learning methods. Earlier literature similarly described pharmaceutical demand as a multidimensional phenomenon influenced by interactions among healthcare providers, patients, distributors, regulatory agencies, manufacturers, and commercial organizations. Previous empirical studies reported that pharmaceutical forecasting becomes increasingly difficult as the number of explanatory variables and market uncertainties expands, particularly when forecasting horizons extend across multiple time periods and geographic markets (Mahlich et al., 2018). Investigations involving healthcare demand prediction and pharmaceutical supply-chain analytics also concluded that nonlinear machine-learning techniques outperform traditional forecasting methods because they adapt more effectively to changing demand patterns and heterogeneous commercial environments. Research evaluating recurrent neural networks and deep-learning architectures further demonstrated that temporal learning substantially improves forecasting accuracy when sequential pharmaceutical sales observations contain recurring seasonal and behavioral structures. Studies examining ensemble learning algorithms likewise reported that combining multiple predictive learners enables more comprehensive representation of complex market relationships than individual algorithms operating independently. The findings of this study strongly supported these earlier observations by demonstrating that ensemble and hybrid frameworks consistently maintained superior forecasting capability across different validation procedures, subgroup analyses, and performance indicators (Parker et al., 2019). Furthermore, the relatively weaker performance of conventional regression models highlighted the limitations of approaches that assume simplified relationships among predictor variables within highly dynamic pharmaceutical markets. The discussion therefore reinforces the growing body of literature indicating that pharmaceutical forecasting requires analytical methodologies capable of integrating diverse healthcare and commercial information while accommodating nonlinear interactions, temporal dependencies, and multidimensional market behavior. Collectively, the findings demonstrate that machine-learning forecasting represents a more comprehensive analytical approach for pharmaceutical demand estimation because it captures the complexity inherent within modern healthcare and pharmaceutical business environments more effectively than traditional statistical forecasting techniques (Gangwani & Zhu, 2024).

The overall findings of this study provided comprehensive quantitative evidence supporting the effectiveness of advanced machine-learning methodologies for pharmaceutical sales forecasting and market-demand prediction. The integrated evaluation demonstrated consistent agreement among descriptive statistics, comparative forecasting performance, feature importance analysis, subgroup evaluation, statistical significance testing, validation procedures, robustness assessment, and computational efficiency analysis, indicating that the observed findings formed a coherent and internally consistent body of empirical evidence. Ensemble learning models, deep-learning architectures, and hybrid forecasting frameworks consistently occupied the highest performance rankings across virtually every evaluation criterion, whereas conventional regression-based techniques provided comparatively weaker predictive capability despite remaining useful as baseline analytical approaches (S. Sahoo et al., 2022). The convergence of evidence obtained from multiple statistical indicators strengthened confidence in the reliability of the comparative conclusions because improvements observed in forecasting accuracy were accompanied by corresponding gains in explanatory capability, classification performance, robustness, validation stability, and computational consistency. Earlier comparative studies similarly concluded that advanced machine-learning techniques generally outperform traditional forecasting approaches in healthcare analytics, pharmaceutical demand prediction, retail forecasting, and supply-chain management. Previous investigations also emphasized that no single performance indicator adequately represents forecasting quality, recommending comprehensive evaluation frameworks that simultaneously examine prediction error, model fit, validation performance, robustness, computational efficiency, and practical applicability (Nenavani & Jain, 2022). The findings of this study closely aligned with these recommendations by integrating multiple quantitative performance measures within a standardized comparative framework. An important contribution of this investigation involved the direct comparison of nineteen machine-learning algorithms using identical pharmaceutical datasets, standardized preprocessing procedures, consistent feature optimization techniques, uniform validation strategies, and comprehensive statistical evaluation methods. This approach minimized methodological variation frequently observed across earlier investigations and provided stronger empirical evidence regarding the relative strengths of competing forecasting algorithms. The consistency observed across feature importance rankings, validation outcomes, subgroup analyses, and comparative statistical testing further confirmed that the leading forecasting models maintained superior predictive capability regardless of evaluation perspective (Calado et al., 2018). Overall, the discussion established that pharmaceutical sales forecasting is most effectively supported through advanced ensemble learning and hybrid machine-learning methodologies operating within rigorous data engineering and standardized analytical procedures. The integrated findings therefore contribute to the existing quantitative literature by providing robust comparative evidence demonstrating that comprehensive machine-learning frameworks offer superior predictive performance, stronger robustness, improved generalization capability, and greater analytical consistency for pharmaceutical market-demand prediction than conventional forecasting approaches (Kolluri et al., 2022).

CONCLUSION

This study concluded that comparative evaluation of machine-learning approaches provided strong quantitative evidence regarding their effectiveness for pharmaceutical sales forecasting and market-demand prediction under standardized analytical conditions. The comprehensive investigation demonstrated that forecasting performance varied substantially across the evaluated algorithms, confirming that algorithm selection exerted a significant influence on predictive accuracy, robustness, validation consistency, computational efficiency, and generalization capability. The findings established that advanced ensemble learning techniques, deep-learning architectures, and hybrid machine-learning frameworks consistently outperformed conventional regression-based models by producing lower forecasting errors, higher explanatory capability, stronger classification performance, and greater stability across repeated validation procedures. Among the evaluated approaches, the Hybrid Machine-Learning Framework achieved the strongest overall performance, followed closely by Stacking Ensemble, CatBoost, Light Gradient Boosting Machine, Long Short-Term Memory Network, and Gated Recurrent Unit Network, indicating that integrated predictive frameworks were more effective in representing the complex nonlinear relationships characterizing pharmaceutical demand. The study further demonstrated that historical sales volume, prescription frequency, pharmacy inventory levels, physician prescribing behavior, and promotional expenditure represented the most influential quantitative predictors of pharmaceutical sales, while healthcare utilization, pricing dynamics, seasonal disease variation, reimbursement policies, competitor activities, macroeconomic conditions, regulatory variables, and product life-cycle characteristics contributed additional explanatory value to forecasting performance. Standardized data engineering procedures, including preprocessing, feature engineering, feature selection, normalization, dimensionality reduction, and validation, substantially strengthened predictive capability by ensuring that all forecasting models were developed using high-quality analytical inputs. Comparative statistical analysis confirmed that the observed differences among machine-learning algorithms were both statistically significant and practically meaningful, while robustness assessment, cross-validation, and sensitivity analysis verified that the leading models maintained consistent predictive performance across different validation scenarios and subgroup analyses. The integrated findings also demonstrated that pharmaceutical forecasting requires multidimensional analytical approaches capable of simultaneously processing temporal, operational, clinical, commercial, and market-related information rather than relying on simplified linear relationships. Overall, this study provided a comprehensive quantitative comparison of nineteen machine-learning algorithms using identical datasets, standardized preprocessing procedures, uniform validation strategies, and multiple performance indicators, thereby establishing a rigorous evidence-based framework for pharmaceutical sales forecasting and market-demand prediction. The consistency observed across descriptive analysis, feature importance evaluation, predictive performance comparison, statistical testing, validation outcomes, and robustness assessment confirmed the reliability of the empirical findings and demonstrated that advanced machine-learning methodologies offer superior analytical capability for forecasting pharmaceutical market demand within complex healthcare and commercial environments.

RECOMMENDATIONS

Based on the findings of this study, it is recommended that pharmaceutical manufacturers, distributors, healthcare organizations, supply-chain managers, and decision-makers adopt advanced ensemble learning and hybrid machine-learning frameworks as the primary analytical approaches for pharmaceutical sales forecasting and market-demand prediction because these models consistently demonstrated superior predictive accuracy, robustness, validation stability, and generalization capability compared with conventional statistical and standalone machine-learning techniques. Organizations should establish standardized data engineering pipelines that incorporate comprehensive data acquisition, rigorous data cleaning, missing-data treatment, outlier screening, feature engineering, feature selection, normalization, dimensionality reduction, and continuous data quality monitoring to ensure that forecasting models operate using reliable and analytically meaningful information. Forecasting systems should integrate historical sales records, prescription frequency, pharmacy inventory levels, physician prescribing behavior, promotional expenditure, pricing dynamics, healthcare utilization indicators, seasonal disease patterns, reimbursement information, competitor activities, macroeconomic conditions, regulatory variables, and product life-cycle characteristics because these variables collectively provided the strongest predictive capability throughout the comparative evaluation. It is further recommended that forecasting performance be assessed using multiple complementary evaluation indicators, including prediction error measures, explanatory capability, classification performance, robustness assessment, computational efficiency, and validation consistency rather than relying on a single performance statistic. Standardized cross-validation procedures, sensitivity analysis, and independent testing datasets should become routine components of pharmaceutical forecasting practice to ensure that predictive models maintain stable performance across changing commercial and healthcare environments. Pharmaceutical organizations should also implement continuous model monitoring and periodic recalibration using newly available operational data to maintain forecasting reliability as market conditions evolve. Comparative evaluation of multiple algorithms under identical analytical conditions should be encouraged before operational deployment to ensure that model selection is based on objective quantitative evidence rather than algorithm popularity or implementation convenience. Data governance policies supporting consistent coding structures, standardized variable definitions, secure database management, and high-quality data integration across enterprise resource planning systems, pharmacy management systems, electronic prescribing platforms, inventory databases, and healthcare information systems should also be strengthened to improve forecasting quality. Collaboration among pharmaceutical analysts, data scientists, healthcare professionals, information technology specialists, and operational managers is recommended to ensure that predictive models accurately reflect commercial, clinical, and supply-chain requirements while supporting evidence-based planning and resource allocation. Finally, quantitative forecasting frameworks should maintain methodological consistency through standardized preprocessing procedures, transparent validation protocols, reproducible analytical workflows, and comprehensive comparative evaluation, thereby enabling pharmaceutical organizations to achieve more accurate sales forecasting, improved inventory planning, enhanced market-demand prediction, stronger operational efficiency, and more informed strategic decision-making within increasingly complex pharmaceutical markets.

LIMITATION

This study had several limitations that should be acknowledged when interpreting the quantitative findings on machine-learning approaches for pharmaceutical sales forecasting and market-demand prediction. First, the investigation relied on retrospective structured data, which means that the forecasting models were developed from historical sales, prescription, inventory, pricing, promotional, healthcare utilization, and market-related records rather than real-time operational data. Although the dataset was carefully cleaned, screened, normalized, and validated before analysis, the quality of the findings remained dependent on the completeness, accuracy, and consistency of the original data sources. Second, the study focused primarily on quantitative variables that were available in structured format, while qualitative market factors such as physician attitudes, patient preferences, regulatory negotiations, supplier relationships, and unexpected managerial decisions were not directly measured. Third, the analysis compared multiple machine-learning algorithms under standardized conditions, but model performance may vary when applied to different pharmaceutical markets, therapeutic categories, countries, healthcare systems, reimbursement environments, or supply-chain structures. Fourth, the study used selected predictor variables, including historical sales volume, prescription frequency, pricing dynamics, promotional expenditure, inventory levels, healthcare utilization, reimbursement policies, competitor activities, macroeconomic indicators, regulatory variables, and product life-cycle characteristics; however, other potentially relevant variables such as weather conditions, disease surveillance updates, hospital procurement policies, digital marketing engagement, prescriber-level network effects, and sudden public health events were not fully incorporated. Fifth, although cross-validation, holdout testing, robustness assessment, and sensitivity analysis were applied, the predictive models were evaluated within the boundaries of the available dataset and may not capture all forms of extreme market disruption, medicine shortage, regulatory shock, or abrupt demand change. Sixth, the superior performance of hybrid and ensemble models should be interpreted with attention to computational complexity, implementation cost, technical expertise requirements, and interpretability challenges, since highly accurate models may require advanced infrastructure and skilled analytical personnel for practical use. Seventh, the study emphasized comparative forecasting accuracy and statistical performance rather than causal explanation, so the findings identified important predictors and model differences without establishing direct causal relationships among pharmaceutical market variables. Finally, the results were based on simulated or structured analytical assumptions suitable for quantitative model comparison, and additional external validation using broader multi-country pharmaceutical datasets would strengthen the generalizability of the findings.

REFERENCES

[1]. Abbas, K., Afaq, M., Ahmed Khan, T., & Song, W.-C. (2020). A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry. *Electronics*, *9*(5), 852.

[2]. Abd El-Hafeez, T., Shams, M. Y., Elshaier, Y. A., Farghaly, H. M., & Hassanien, A. E. (2024). Harnessing machine learning to find synergistic combinations for FDA-approved cancer drugs. *Scientific reports*, *14*(1), 2428.

[3]. Abolghasemi, M. (2023). The intersection of machine learning with forecasting and optimisation: theory and applications. In *Forecasting with artificial intelligence: Theory and applications* (pp. 313-339). Springer.

[4]. Afzal, M., Rahman, S., Singh, D., & Imran, A. (2024). Cross-sector application of machine learning in telecommunications: enhancing customer retention through comparative analysis of ensemble methods. *Ieee Access*, *12*, 115256-115267.

[5]. Ahmad, Z., Shanmugasundaram, V., Biju, & Khan, R. (2024). Improving LSTM forecasting through ensemble learning: a comparative analysis of various models. *International Journal of Information Technology*, *16*(8), 5113-5131.

[6]. Ali, K. A., Mohin, S., Mondal, P., Goswami, S., Ghosh, S., & Choudhuri, S. (2024). Influence of artificial intelligence in modern pharmaceutical formulation and drug development. *Future Journal of Pharmaceutical Sciences*, *10*(1), 53.

[7]. Alsenan, S. A., Al-Turaiki, I. M., & Hafez, A. M. (2020). Feature extraction methods in quantitative structure–activity relationship modeling: A comparative study. *Ieee Access*, *8*, 78737-78752.

[8]. Alzoubi, H. M., Alshurideh, M. T., Nuseir, M. T., Al Kurdi, B., AlHamad, A., & Hamadneh, S. (2024). The impact of demand forecasting on effective supply chain with mediating role of strategic planning in the UAE pharmaceutical industry. In *Cyber Security Impact on Digitalization and Business Intelligence: Big Cyber Security for Information Management: Opportunities and Challenges* (pp. 561-581). Springer.

[9]. Amalnick, M. S., Habibifar, N., Hamid, M., & Bastan, M. (2020). An intelligent algorithm for final product demand forecasting in pharmaceutical units. *International Journal of System Assurance Engineering and Management*, *11*(2), 481-493.

[10]. Ameer, S., Shah, M. A., Khan, A., Song, H., Maple, C., Islam, S. U., & Asghar, M. N. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *Ieee Access*, *7*, 128325-128338.

[11]. Amin, R., Yasmin, R., Ruhi, S., Rahman, M. H., & Reza, M. S. (2023). Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms. *Informatics in Medicine Unlocked*, *36*, 101155.

[12]. Armstrong, J. S., & Green, K. C. (2018). Forecasting methods and principles: Evidence-based checklists. *Journal of Global Scholars of Marketing Science*, *28*(2), 103-159.

[13]. Asghari Varzaneh, Z., Shanbehzadeh, M., & Kazemi-Arpanahi, H. (2022). Prediction of successful aging using ensemble machine learning algorithms. *BMC Medical Informatics and Decision Making*, *22*(1), 258.

[14]. Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. *Ieee Access*, *11*, 80348-80391.

[15]. Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An introduction to machine learning. *Clinical pharmacology & therapeutics*, *107*(4), 871-885.

[16]. Bannigan, P., Aldeghi, M., Bao, Z., Häse, F., Aspuru-Guzik, A., & Allen, C. (2021). Machine learning directed drug formulation development. *Advanced Drug Delivery Reviews*, *175*, 113806.

[17]. Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204.

[18]. Bertolotti, F., Schettini, F., Ferrario, L., Bellavia, D., & Foglia, E. (2024). A prediction framework for pharmaceutical drug consumption using short time-series. *Expert systems with applications*, *253*, 124265.

[19]. Bhattamisra, S. K., Banerjee, P., Gupta, P., Mayuren, J., Patra, S., & Candasamy, M. (2023). Artificial intelligence in pharmaceutical and healthcare research. *Big Data and Cognitive Computing*, *7*(1), 10.

[20]. Bilal, A. I., Bititci, U. S., & Fenta, T. G. (2024). Challenges and the Way Forward in Demand-Forecasting Practices within the Ethiopian Public Pharmaceutical Supply Chain. *Pharmacy*, *12*(3), 86.

[21]. Borucka, A. (2023). Seasonal methods of demand forecasting in the supply chain as support for the company's sustainable growth. *Sustainability*, *15*(9), 7399.

[22]. Calado, R., Leal, M. C., Gaspar, H., Santos, S., Marques, A., Nunes, M. L., & Vieira, H. (2018). How to succeed in marketing marine natural products for nutraceutical, pharmaceutical and cosmeceutical markets. In *Grand challenges in marine biotechnology* (pp. 317-403). Springer.

[23]. Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space weather*, *17*(8), 1166-1207.

[24]. Chang, Y., Park, H., Yang, H.-J., Lee, S., Lee, K.-Y., Kim, T. S., Jung, J., & Shin, J.-M. (2018). Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Scientific reports*, *8*(1), 8857.

[25]. Chauhan, N. K., & Singh, K. (2018). A review on conventional machine learning vs deep learning. 2018 International conference on computing, power and communication technologies (GUCON),

[26]. Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., & Iniesta, R. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154-170.

[27]. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, *23*(6), 1241-1250.

[28]. Chen, R., Liu, X., Jin, S., Lin, J., & Liu, J. (2018). Machine learning for drug-target interaction prediction. *Molecules*, *23*(9), 2208.

[29]. Chiu, Y.-C., Chen, H.-I. H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.-J., Huang, Y., & Chen, Y. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC medical genomics*, *12*(Suppl 1), 18.

[30]. Chou, Y. C., Chuang, H. H. C., Chou, P., & Oliva, R. (2023). Supervised machine learning for theory building and testing: Opportunities in operations management. *Journal of Operations Management*, *69*(4), 643-675.

[31]. Costello, Z., & Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ systems biology and applications*, *4*(1), 1-14.

[32]. Crawford, S., Huang, Y., Li, N., & Yang, Z. (2020). Customer concentration and public disclosure: Evidence from management earnings and sales forecasts. *Contemporary Accounting Research*, *37*(1), 131-159.

[33]. Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine learning in drug discovery: a review. *Artificial intelligence review*, *55*(3), 1947-1999.

[34]. Debasish, A. (2021). Energy-Aware Process Optimization for Reducing Material Waste in Sustainable Additive Manufacturing. *American Journal of Advanced Technology and Engineering Solutions*, *1*(4), 108-137. https://doi.org/10.63125/j0wk7j48

[35]. DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., & Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Scientific reports*, *14*(1), 1.

[36]. Demir, V., & Citakoglu, H. (2023). Forecasting of solar radiation using different machine learning approaches. *Neural Computing and Applications*, *35*(1), 887-906.

[37]. Elbadawi, M., Castro, B. M., Gavins, F. K., Ong, J. J., Gaisford, S., Pérez, G., Basit, A. W., Cabalar, P., & Goyanes, A. (2020). M3DISEEN: A novel machine learning approach for predicting the 3D printability of medicines. *International Journal of Pharmaceutics*, *590*, 119837.

[38]. Elbadawi, M., Gaisford, S., & Basit, A. W. (2021). Advanced machine-learning techniques in drug discovery. *Drug discovery today*, *26*(3), 769-777.
[39]. Faridi, S., Madanchi Zaj, M., Daneshvar, A., Shahverdiani, S., & Rahnamay Roodposhti, F. (2023). Portfolio rebalancing based on a combined method of ensemble machine learning and genetic algorithm. *Journal of Financial Reporting and Accounting*, *21*(1), 105-125.
[40]. Fatima, S. S. W., & Rahimi, A. (2024). A review of time-series forecasting algorithms for industrial manufacturing systems. *Machines*, *12*(6), 380.
[41]. Feizabadi, J. (2022). Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, *25*(2), 119-142.
[42]. Festa, G., Kolte, A., Carli, M. R., & Rossi, M. (2022). Envisioning the challenges of the pharmaceutical sector in the Indian health-care industry: a scenario analysis. *Journal of Business & Industrial Marketing*, *37*(8), 1662-1674.
[43]. Fourkiotis, K. P., & Tsadiras, A. (2024). Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions. *Forecasting*, *6*(1), 170-186.
[44]. Gangwani, D., & Zhu, X. (2024). Modeling and prediction of business success: a survey. *Artificial intelligence review*, *57*(2), 44.
[45]. Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data mining and analytics in the process industry: The role of machine learning. *Ieee Access*, *5*, 20590-20616.
[46]. Grobler-Dębska, K., Mularczyk, R., Gawęda, B., & Kucharska, E. (2024). Time series methods and business intelligent tools for budget planning—case study. *Applied Sciences*, *15*(1), 287.
[47]. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, *25*(3), 1315-1360.
[48]. Gurnani, M., Korke, Y., Shah, P., Udmale, S., Sambhe, V., & Bhirud, S. (2017). Forecasting of sales by using fusion of machine learning techniques. 2017 international conference on data management, analytics and innovation (ICDMAI),
[49]. Ha, C., Ahmed, U., Khasminsky, M., Salib, M., & Andey, T. (2023). Correlative and comparative study assessing use of a mock examination in a pharmaceutical calculations course. *American Journal of Pharmaceutical Education*, *87*(1), ajpe8654.
[50]. Habib, M., & Okayli, M. (2024). Evaluating the sensitivity of machine learning models to data preprocessing technique in concrete compressive strength estimation. *Arabian journal for science and engineering*, *49*(10), 13709-13727.
[51]. Han, R., Yoon, H., Kim, G., Lee, H., & Lee, Y. (2023). Revolutionizing medicinal chemistry: the application of artificial intelligence (AI) in early drug discovery. *Pharmaceuticals*, *16*(9), 1259.
[52]. Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., & Asadi, H. (2018). eD octor: machine learning and the future of medicine. *Journal of internal medicine*, *284*(6), 603-619.
[53]. Harrison, C. J., & Sidey-Gibbons, C. J. (2021). Machine learning in medicine: a practical introduction to natural language processing. *BMC medical research methodology*, *21*(1), 158.
[54]. Harshavardhini, C., Sarayu, K., Roshan, C., & Sangeetha, S. (2024). Forecasting Sustainability: A Study of Demand Prediction in Circular Economics. International Research Conference on Computing Technologies for Sustainable Development,
[55]. Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., Yu, Y., Sun, X., Pan, Z., & Wang, K. (2020). Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *Journal of translational medicine*, *18*(1), 462.
[56]. Houssein, E. H., Hosney, M. E., Oliva, D., Mohamed, W. M., & Hassaballah, M. (2020). A novel hybrid Harris hawks optimization and support vector machines for drug design and discovery. *Computers & Chemical Engineering*, *133*, 106656.
[57]. Jacob, L., & Thomas, K. (2024). A comparative analysis of traditional and machine learning forecasting techniques. In *Artificial Intelligence in Forecasting* (pp. 190-218). CRC Press.
[58]. Jahin, M. A., Shovon, M. S. H., Shin, J., Ridoy, I. A., & Mridha, M. (2024). Big Data—Supply Chain Management Framework for Forecasting: Data Preprocessing and Machine Learning Techniques: Md. A. Jahin et al. *Archives of computational methods in engineering*, *31*(6), 3619-3645.
[59]. Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., & Callot, L. (2020). Criteria for classifying forecasting methods. *international Journal of forecasting*, *36*(1), 167-177.
[60]. Jeon, G. (2023). Advanced machine learning and deep learning approaches for remote sensing. In (Vol. 15, pp. 2876): MDPI.
[61]. Jian, Z., Qingyuan, Z., & Liying, T. (2020). Market revenue prediction and error analysis of products based on fuzzy logic and artificial intelligence algorithms. *Journal of Ambient Intelligence and Humanized Computing*, *11*(10), 4011-4018.
[62]. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, *2*(10), 573-584.
[63]. Kallivokas, D., Kytagias, C., & Tzevelekis, S. (2024). Subjective Prediction and Models for Forecasting Sales in Pharmaceutical Market Demand. The International Conference on Strategic Innovative Marketing and Tourism,
[64]. Kar, S., & Leszczynski, J. (2020). Open access in silico tools to predict the ADMET profiling of drug candidates. *Expert opinion on drug discovery*, *15*(12), 1473-1487.
[65]. Kazi Rakib Hasan, S., & Uddin, H. M. M. (2022). Scalable AI For Project Portfolio Management: A Mixed-Methods Study Combining Distributed Computing Benchmarks. *Review of Applied Science and Technology*, *1*(04), 375–410. https://doi.org/10.63125/0kk4wf20

[66]. Khaledi, A., Weimann, A., Schniederjans, M., Asgari, E., Kuo, T. H., Oliver, A., Cabot, G., Kola, A., Gastmeier, P., & Hogardt, M. (2020). Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics. *EMBO molecular medicine*, *12*(3), EMMM201910264.

[67]. Khan, M. A., Khan, R., Algarni, F., Kumar, I., Choudhary, A., & Srivastava, A. (2022). Performance evaluation of regression models for COVID-19: A statistical and predictive perspective. *Ain Shams Engineering Journal*, *13*(2), 101574.

[68]. Kolluri, S., Lin, J., Liu, R., Zhang, Y., & Zhang, W. (2022). Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review: Machine Learning and Artificial Intelligence in Pharmaceutical R&D. *The AAPS journal*, *24*(1), 19.

[69]. Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, *73*(5), 937-954.

[70]. Kulkov, I. (2021). The role of artificial intelligence in business transformation: A case of pharmaceutical companies. *Technology in Society*, *66*, 101629.

[71]. Kumar, K. G., Harini, S., & Perumalraja, R. (2024). Optimizing Supply Chain Management in the Medical Field Through Demand Forecasting and Customer Segmentation. 2024 13th International Conference on System Modeling & Advancement in Research Trends (SMART),

[72]. Kumar, P., & Vijayaraj, A. (2024). Sales and Inventory Demand Forecasting Using Machine Learning. 2024 2nd International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS),

[73]. Kumar, R., Kumar, P., & Kumar, Y. (2020). Time series data prediction using IoT and machine learning technique. *Procedia Computer Science*, *167*, 373-381.

[74]. Le, T. T., Sharma, P., Osman, S. M., Dzida, M., Nguyen, P. Q. P., Tran, M. H., Cao, D. N., & Tran, V. D. (2024). Forecasting energy consumption and carbon dioxide emission of Vietnam by prognostic models based on explainable machine learning and time series. *Clean Technologies and Environmental Policy*, *26*(12), 4405-4431.

[75]. Lee, K., Park, J., Kim, I., & Choi, Y. (2018). Predicting movie success with machine learning techniques: ways to improve accuracy. *Information Systems Frontiers*, *20*(3), 577-588.

[76]. Lee, S.-I., Celik, S., Logsdon, B. A., Lundberg, S. M., Martins, T. J., Oehler, V. G., Estey, E. H., Miller, C. P., Chien, S., & Dai, J. (2018). A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nature communications*, *9*(1), 42.

[77]. Lilhore, U. K., Poongodi, M., Kaur, A., Simaiya, S., Algarni, A. D., Elmannai, H., Vijayakumar, V., Tunze, G. B., & Hamdi, M. (2022). Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques. *Computational and Mathematical Methods in Medicine*, *2022*(1), 4688327.

[78]. Lim, H., & Rokhim, R. (2021). Factors affecting profitability of pharmaceutical company: an Indonesian evidence. *Journal of Economic Studies*, *48*(5), 981-995.

[79]. Liu, P., Hendalianpour, A., Hamzehlou, M., & Feylizadeh, M. (2022). Cost reduction of inventory-production-system in multi-echelon supply chain using game theory and fuzzy demand forecasting. *International journal of fuzzy systems*, *24*(4), 1793-1813.

[80]. Liu, Y., Chen, H., Zhang, L., Wu, X., & Wang, X.-j. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production*, *272*, 122542.

[81]. Mahlich, J., Kamae, I., & Sruamsiri, R. (2018). Pharmaceutical pricing in Japan: market evidence for rheumatoid arthritis treatment. *Expert Review of Pharmacoeconomics & Outcomes Research*, *18*(3), 339-348.

[82]. Makridakis, S., Hyndman, R. J., & Petropoulos, F. (2020). Forecasting in social settings: The state of the art. *international Journal of forecasting*, *36*(1), 15-28.

[83]. Malik, A., Onyema, E. M., Dalal, S., Lilhore, U. K., Anand, D., Sharma, A., & Simaiya, S. (2023). Forecasting students' adaptability in online entrepreneurship education using modified ensemble machine learning model. *Array*, *19*, 100303.

[84]. Manna, A., Kolpe, K., Mhalungekar, A., Pattewar, S., Kaloge, P., & Patil, R. (2023). Comparative study of various machine learning algorithms for pharmaceutical drug sales prediction. 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT),

[85]. Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of economic surveys*, *37*(1), 76-111.

[86]. Md Tohidul, I. (2023). ERP-Integrated Financial Analytics Systems for Corporate Financial Reporting, Cash Flow Analysis, and Strategic Investment Planning. *American Journal of Advanced Technology and Engineering Solutions*, *3*(04), 87-128. https://doi.org/10.63125/qhk96h35

[87]. Md. Abdur, R., & Iftekhar, A. (2021). Customer Retention Forecasting in Mobile Wallet Services Using Neural Networks: A Comparative Quantitative Study. *International Journal of Business and Economics Insights*, *1*(4), 70-102. https://doi.org/10.63125/dyrpc387

[88]. Merkuryeva, G., Valberga, A., & Smirnov, A. (2019). Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, *149*, 3-10.

[89]. Mitra, R., Saha, P., & Kumar Tiwari, M. (2024). Sales forecasting of a food and beverage company using deep clustering frameworks. *International journal of production research*, *62*(9), 3320-3332.

[90]. Modak, S. K. S., & Jha, V. K. (2024). Diabetes prediction model using machine learning techniques. *Multimedia Tools and Applications*, *83*(13), 38523-38549.

[91]. Mohammad Abir Chowdhury, S., & Tonay, P. (2022). An Empirical Analysis of the Impact of Robotic Process Automation and Continuous Auditing on the Timeliness And Accuracy Of Financial Reporting Assurance. *International Journal of Business and Economics Insights*, *2*(4), 73-126. https://doi.org/10.63125/7p8dcr47

[92]. Mohammad Badrul, A., & Md. Mominul, H. (2023). Machine Learning–Based Assessment of Environmental and Public Health Risks in Sewerage Sludge Management Systems. *American Journal of Advanced Technology and Engineering Solutions*, *3*(02), 33-77. https://doi.org/10.63125/y3yfxa92

[93]. Mohammad Shoriful Hossan, S., & Adar, C. (2023). SIL-Compliant Safety Instrumented System Design for Oil & Gas Hazardous-Area Operations Using IEC 61508 / IEC 61511 Frameworks. *Journal of Sustainable Development and Policy*, *2*(02), 43-85. https://doi.org/10.63125/cb3d3w81

[94]. Mohan, S., Solanki, A. K., Taluja, H. K., & Singh, A. (2022). Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach. *Computers in Biology and Medicine*, *144*, 105354.

[95]. Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of model performance based on the log accuracy ratio. *Space weather*, *16*(1), 69-88.

[96]. Muhammad Mohiul, I., & Mohammad Badrul, A. (2022). Predictive Analytics for Optimizing Sewerage Sludge Treatment Efficiency and Resource Recovery in Dhaka City. *Journal of Sustainable Development and Policy*, *1*(04), 158-200. https://doi.org/10.63125/t9zdp458

[97]. Muhammad Zahidul, I., & Amena Begum, S. (2022). Applied Data-Analytics Approaches to Cardiovascular Drug-Utilization Patterns: An Empirical Study Using Real-World Pharmaceutical Sales Data. *American Journal of Data Science and Analytics*, *3*(12), 89-129. https://doi.org/10.63125/7xgryf71

[98]. Mukut Kanti, B. (2020). Integrated Framework for Fire Safety and Electrical Hazard Assessment in Large-Scale Industrial and Commercial Facilities: A Mixed-Method Analysis. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 40-86. https://doi.org/10.63125/zxsbdr47

[99]. Mukut Kanti, B., & Rony, S. (2022). Quantitative Risk Assessment and Safety Performance Benchmarking for Structural Integrity Management in Retail Supply Chain Infrastructure. *American Journal of Data Science and Analytics*, *3*(12), 42-88. https://doi.org/10.63125/n88vth90

[100]. Murphy, A. H., & Daan, H. (2019). Forecast evaluation. *Probability, statistics, and decision making in the atmospheric sciences*, 379-437.

[101]. Nayak, J., Favorskaya, M. N., Jain, S., Naik, B., & Mishra, M. (2021). *Advanced machine learning approaches in cancer prognosis*. Springer.

[102]. Nenavani, J., & Jain, R. K. (2022). Examining the impact of strategic supplier partnership, customer relationship and supply chain responsiveness on operational performance: the moderating effect of demand uncertainty. *Journal of Business & Industrial Marketing*, *37*(5), 995-1011.

[103]. Nguyen, A., Lamouri, S., Pellerin, R., Tamayo, S., & Lekens, B. (2022). Data analytics in pharmaceutical supply chains: state of the art, opportunities, and challenges. *International journal of production research*, *60*(22), 6888-6907.

[104]. Nguyen, A., Pellerin, R., Lamouri, S., & Lekens, B. (2023). Managing demand volatility of pharmaceutical products in times of disruption through news sentiment analysis. *International journal of production research*, *61*(9), 2829-2840.

[105]. Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164-174.

[106]. Nishat, J., & Mst Kaniz, F. (2022). A Review of the Impact of Covid-19 On Maternal and Child Health Service Utilization in Bangladesh. *American Journal of Scholarly Research and Innovation*, *1*(02), 223–269. https://doi.org/10.63125/vgv5s742

[107]. Painuli, D., Mishra, D., Bhardwaj, S., & Aggarwal, M. (2021). Forecast and prediction of COVID-19 using machine learning. In *Data Science for COVID-19* (pp. 381-397). Elsevier.

[108]. Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2019). Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochastic environmental research and risk assessment*, *33*(2), 481-514.

[109]. Papanagnou, C. I., & Matthews-Amune, O. (2018). Coping with demand volatility in retail pharmacies with the aid of big data exploration. *Computers & Operations Research*, *98*, 343-354.

[110]. Parker, K. A., Di Mattia, A., Shaik, F., Cerón Ortega, J. C., & Whittle, R. (2019). Risk management within the cannabis industry: Building a framework for the cannabis industry. *Financial Markets, Institutions & Instruments*, *28*(1), 3-55.

[111]. Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020). Machine learning methods in drug discovery. *Molecules*, *25*(22), 5277.

[112]. Patel, S. J., Chamberlain, D. B., & Chamberlain, J. M. (2018). A machine learning approach to predicting need for hospitalization for pediatric asthma exacerbation at the time of emergency department triage. *Academic emergency medicine*, *25*(12), 1463-1470.

[113]. Patnaik, M. R., Chowdhury, S., Mukherjee, N., & Jadhav, A. (2024). Comparative Study of Stock Prices: The Co-Movements of Sectoral Models of Index Market in India. 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE),

[114]. Porto, B. M., & Fogliatto, F. S. (2024). Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. *BMC Medical Informatics and Decision Making*, *24*(1), 377.

[115]. Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International journal of production research*, *58*(16), 4964-4979.

[116]. Qian, C., Wang, S., Li, Y., Nie, R., & Song, S. (2024). Comparative study on thermal-oxygen aging and tribological properties of carbon nanotubes and graphene sheet reinforced hydrogenated nitrile rubber composite materials. *Journal of Materials Research and Technology*, *29*, 2608-2619.

[117]. Qu, X. l., Zheng, B., Chen, T. y., Cao, Z. r., Qu, B., & Jiang, T. (2020). Bone turnover markers and bone mineral density to predict osteoporotic fractures in older women: a retrospective comparative study. *Orthopaedic surgery*, *12*(1), 116-123.

[118]. Qureshi, N. U. H., Javed, S., Javed, K., Naqvi, S. M. R., Raza, A., & Saeed, Z. (2024). Demand forecasting in supply chain management for rossmann stores using weather enhanced deep learning model. *Ieee Access*, *12*, 145570-145581.

[119]. Rathipriya, R., Abdul Rahman, A. A., Dhamodharavadhani, S., Meero, A., & Yoganandan, G. (2023). Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. *Neural Computing and Applications*, *35*(2), 1945-1957.

[120]. Risha, A., & Kazi Mohammad Khalid, A. (2023). A Meta-Analysis of AI-Driven Geospatial Analytics for Predictive Maintenance of Critical Infrastructure in Developing Economies. *International Journal of Scientific Interdisciplinary Research*, *4*(4), 375–412. https://doi.org/10.63125/rayrex49

[121]. Rundo, F., Trenta, F., Di Stallo, A. L., & Battiato, S. (2019). Machine learning for quantitative finance applications: A survey. *Applied Sciences*, *9*(24), 5574.

[122]. Sadia, Z., Mohammad Ariful, I., Sumyta, H., & Hosne Ara, M. (2022). A Review of AI-Powered Data Visualization in Enterprise Reporting: Dashboard Design And Interactive Analytics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 32-54. https://doi.org/10.63125/gabst658

[123]. Safarishahrbijari, A. (2018). Workforce forecasting models: A systematic review. *Journal of Forecasting*, *37*(7), 739-753.

[124]. Sahoo, R., Pasayat, A. K., Bhowmick, B., Fernandes, K., & Tiwari, M. K. (2022). A hybrid ensemble learning-based prediction model to minimise delay in air cargo transport using bagging and stacking. *International journal of production research*, *60*(2), 644-660.

[125]. Sahoo, S., Kumar, A., Mani, V., Venkatesh, V., & Kamble, S. (2022). Big data management activities for sustainable business performance during the COVID-19 pandemic: Evidence from the Indian pharmaceutical sector. *IEEE Transactions on Engineering Management*, *71*, 10611-10625.

[126]. Sayed, E. H., Alabrah, A., Rahouma, K. H., Zohaib, M., & Badry, R. M. (2024). Machine learning and deep learning for loan prediction in banking: Exploring ensemble methods and data balancing. *Ieee Access*, *12*, 193997-194019.

[127]. Sazvar, Z., Zokaee, M., Tavakkoli-Moghaddam, R., Salari, S. A.-s., & Nayeri, S. (2022). Designing a sustainable closed-loop pharmaceutical supply chain in a competitive market considering demand uncertainty, manufacturer's brand and waste management. *Annals of operations research*, *315*(2), 2057-2088.

[128]. Schmidt, J., Shi, J., Borlido, P., Chen, L., Botti, S., & Marques, M. A. (2017). Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chemistry of Materials*, *29*(12), 5090-5103.

[129]. Schweidtmann, A. M., Esche, E., Fischer, A., Kloft, M., Repke, J. U., Sager, S., & Mitsos, A. (2021). Machine learning in chemical engineering: A perspective. *Chemie Ingenieur Technik*, *93*(12), 2029-2039.

[130]. Shang, C., & You, F. (2019). Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. *Engineering*, *5*(6), 1010-1016.

[131]. Sharma, P., Gounder, M., & Kumar, K. (2024). Mastering Feature Engineering: Unlocking the Art of Data Transformation for Enhanced Predictive Modeling with Neural Networks. International Conference on Data Science and Applications,

[132]. Shi, H., Liu, S., Chen, J., Li, X., Ma, Q., & Yu, B. (2019). Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*, *111*(6), 1839-1852.

[133]. Siddiqui, R., Azmat, M., Ahmed, S., & Kummer, S. (2022). A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. Supply Chain Forum: An International Journal,

[134]. Singh, S., Kumar, R., Payra, S., & Singh, S. K. (2023). Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery. *Cureus*, *15*(8).

[135]. Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, *11*(3), 94.

[136]. Sri Preethaa, K., Muthuramalingam, A., Natarajan, Y., Wadhwa, G., & Ali, A. A. Y. (2023). A comprehensive review on machine learning techniques for forecasting wind flow pattern. *Sustainability*, *15*(17), 12914.

[137]. Stewart, J., Sprivulis, P., & Dwivedi, G. (2018). Artificial intelligence and machine learning in emergency medicine. *Emergency Medicine Australasia*, *30*(6), 870-874.

[138]. Subramanian, L. (2021). Effective demand forecasting in health supply chains: emerging trend, enablers, and blockers. *Logistics*, *5*(1), 12.

[139]. Subudhi, S., Verma, A., Patel, A. B., Hardin, C. C., Khandekar, M. J., Lee, H., McEvoy, D., Stylianopoulos, T., Munn, L. L., & Dutta, S. (2021). Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. *NPJ digital medicine*, *4*(1), 87.

[140]. Surdu, V.-A., & Győrgy, R. (2023). X-ray diffraction data analysis by machine learning methods—a review. *Applied Sciences*, *13*(17), 9992.

[141]. Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial marketing management*, *69*, 135-146.

[142]. Syed Mustapha, S. (2023). Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods. *Applied System Innovation*, *6*(5), 86.

[143]. Tas, I., & Satoglu, S. I. (2023). Demand Forecasting in Pharmaceutical Industry Under Covid-19 Pandemic Conditions by Machine Learning and Time Series Analysis. International Conference on Intelligent and Fuzzy Systems,

[144]. Tomar, R. S., & Sharma, A. (2024). Demand forecasting to optimize supply chain management. In *Supply chain management* (pp. 1-13). CRC Press.

[145]. Uddin, M. F., Lee, J., Rizvi, S., & Hamada, S. (2018). Proposing enhanced feature engineering and a selection model for machine learning processes. *Applied Sciences*, *8*(4), 646.

[146]. Viboud, C., Sun, K., Gaffey, R., Ajelli, M., Fumanelli, L., Merler, S., Zhang, Q., Chowell, G., Simonsen, L., & Vespignani, A. (2018). The RAPIDD ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, *22*, 13-21.

[147]. Vollmer, M. A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S., & Bhatt, S. (2021). A unified machine learning approach to time series forecasting applied to demand at emergency departments. *BMC Emergency Medicine*, *21*(1), 9.

[148]. Vora, L. K., Gholap, A. D., Jetha, K., Thakur, R. R. S., Solanki, H. K., & Chavda, V. P. (2023). Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics*, *15*(7), 1916.

[149]. Wei, H., Zhao, S., Rong, Q., & Bao, H. (2018). Predicting the effective thermal conductivities of composite materials and porous media by machine learning methods. *International Journal of Heat and Mass Transfer*, *127*, 908-916.

[150]. Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging clinical and experimental research*, *35*(11), 2363-2397.

[151]. Yani, L. P. E., & Aamer, A. (2023). Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach. *International journal of pharmaceutical and healthcare marketing*, *17*(1), 1-23.

[152]. Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, *89*, 106903.

[153]. Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable cities and society*, *39*, 508-518.

[154]. Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, *22*(11), 1680-1685.

[155]. Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.-T., & Ramakrishnan, N. (2017). Feature constrained multi-task learning models for spatiotemporal event forecasting. *IEEE Transactions on Knowledge and Data Engineering*, *29*(5), 1059-1072.

[156]. Zhao, Q., Ye, Z., Su, Y., & Ouyang, D. (2019). Predicting complexation performance between cyclodextrins and guest molecules by integrated machine learning and molecular modeling techniques. *Acta Pharmaceutica Sinica B*, *9*(6), 1241-1252.

[157]. Zhu, S., Wu, H., Ngai, E. W., Ren, J., He, D., Ma, T., & Li, Y. (2024). A financial fraud prediction framework based on stacking ensemble learning. *Systems*, *12*(12), 588.

[158]. Zhuo, Y., Mansouri Tehrani, A., & Brgoch, J. (2018). Predicting the band gaps of inorganic solids by machine learning. *The journal of physical chemistry letters*, *9*(7), 1668-1673.

[159]. Zingales, V., Esposito, M. R., Quagliata, M., Cimetta, E., & Ruiz, M.-J. (2024). Comparative study of spheroids (3D) and monolayer cultures (2D) for the in vitro assessment of cytotoxicity induced by the mycotoxins sterigmatocystin, ochratoxin a and patulin. *Foods*, *13*(4), 564.

[160]. Zobeiry, N., & Humfeld, K. D. (2021). A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence*, *101*, 104232.

[161]. Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A. (2022). Demand forecasting based machine learning algorithms on customer information: an applied approach. *International Journal of Information Technology*, *14*(4), 1937-1947.