International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 7 Issue 2 PP 157-164, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I2P120

Original Article

Data-Driven Pharmaceutical Sales Analysis and Prediction in the Healthcare Industry

Dr. Pradeep Laxkar
Associate Professor Department of Computer Science and Engineering, ITM (SLS) University, Vadodara, Gujarat.

Received On: 22/03/2026 ***Revised On: 21/04/2026*** ***Accepted On: 29/04/2026*** ***Published On: 05/05/2026***

Abstract - *Forecasting pharmaceutical drug sales is a significant challenge for healthcare organizations and pharmaceutical companies due to factors such as seasonality, weather conditions, local health crises, import issues, currency fluctuations, and economic instability. These factors can contribute to shortages of drugs or a surplus of drugs, impacting both drug availability and drug operations. The challenges discussed above have been addressed in this study by presenting an efficient model for Pharmaceutical Sales Analysis and Forecasting using a combination of Hybrid Random Forest (RF) and Gated Recurrent Unit (RF+GRU) model on the Pharmaceutical sales dataset available from Kaggle. The data was preprocessed using techniques like outlier detection, handling missing data, label encoding, normalization, feature extraction, and data balancing using SMOTE to improve data quality and predictive accuracy. The hybrid model proposed is built by combining the best of both worlds from the feature extraction capabilities of RF and the sequential learning capability of GRU in order to increase forecast accuracy and minimize forecast error. A 96.1% accuracy (ACC) rate, 96.8% precision (PRE), 96% recall (REC), and 96.4% F1-score (F1) were attained by the suggested RF+GRU model in comparison to the other models, including XGBoost, SVM, and KNN, according to the experimental data. The results demonstrate that the hybrid forecasting model is reliable and effective for predicting sales for the pharmaceutical industry, and can be utilized to support the inventory management system and healthcare decision-making system.*

Keywords - *Sales Forecasting, Machine Learning, Time Series Analysis, Pharmaceutical Industry, Seasonality Effects.*

1. Introduction

The health care industry is an important contributor to enhancing human life through the provision, medical services and patient care [1]. The effective prediction of drug demand and pharmaceuticals sales is vital to ensuring adequate stock levels, minimizing operational costs and enabling efficient health services provision [2]. But, demand for medicines is uncertain, patient demands are shifting and the consumption of medicines is fluctuating, all of which pose significant challenges in the pharmaceutical supply chain management [3]. Poor forecasting can result in overstocking or understocking a healthcare facility, waste of medicines, and losses for manufacturers [4][5].

Historically, pharmaceuticals sales forecasting has been based on statistical and mathematical methods, which were based on the historical sales data, prescription patterns, patient information, and patterns of healthcare utilization [6][7]. These approaches were valuable for inventory planning and resource allocation, but failed frequently to cope with complex and volatile market situations. Conventional forecasting methods were further hindered by external factors like economic fluctuations, public health crises, regulatory shifts, and consumer behavior, in health systems that operate in dynamic environments [8][9].

As the amount of digital healthcare data continues to expand, Machine Learning (ML) is proving to be a valuable tool for pharmaceutical sales analyses and forecasting [10][11]. With ML techniques [12], it is possible to process large-scale datasets efficiently and discover patterns and relationships that may not be apparent in the data [13][14]. These are the approaches that help medical organizations and pharmaceutical firms to make more accurate forecasts, manage inventory more effectively, lower medicine waste and make better decisions. Predictive analytics has transformed pharmaceutical forecasting from an intuitive process to a more evidence-based and data-driven one [15][16].

In recent years, predictive capabilities have been further improved in pharmaceutical forecasting applications by improvements made in the field of AI and DL [17]. Adaptive forecasting in dynamic environments is possible with AI-based systems that can continuously learn from new healthcare and sales data. Pharmaceutical sales data is highly complex and contains non-linear relationships that are well captured by DL techniques [18]. As a result, the pharmaceutical business is increasingly relying on AI and DL for reliable sales forecasting, streamlined inventory management, and environmentally friendly healthcare operations [19][20][21][22].

1.1. Motivation and Contribution

The motivation behind the study is the increasing need for accurate pharmaceutical sales forecasts, particularly in the context of guaranteeing the availability of medications, reducing wastage, and improving the efficiency of the healthcare supply chain. Traditional forecasting methods can prove to be challenging when it comes to predicting the complex and dynamic nature of pharmaceutical sales trends,

where patient needs and market dynamics are constantly changing. As the amount of health information continues to grow and ML and DL techniques advance, new opportunities emerge to make more accurate predictions and decisions. Therefore, the present work is to develop an efficient hybrid forecasting model to enhance the analysis of pharmaceutical sales and assist the pharmaceutical inventory management in health care systems in an effective manner. This study has significantly contributed in the following ways:

- Utilized a large-scale Pharma sales dataset collected from Kaggle containing pharmaceutical sales records for model training and evaluation.
- Used data preparation methods to enhance data consistency and quality, including treating missing values, removing outliers, encoding labels, and normalizing.
- Performed feature extraction and SMOTE-based data balancing to enhance prediction performance and handle class imbalance in the Pharma sales dataset.
- Proposed a Hybrid RF+GRU model that combines feature extraction and sequential learning capabilities for effective pharmaceutical sales forecasting.
- Verified the efficacy and dependability of the forecasting system by assessing the suggested model using conventional performance measures including REC, ACC, PRE, and F1.

The justification of this study lies in the growing need for accurate pharmaceutical sales forecasting to reduce drug shortages, minimize overstocking, and improve healthcare supply chain management. Existing forecasting methods often face limitations in handling complex and sequential pharmaceutical sales data, which affects prediction ACC and decision-making efficiency. The novelty of this work is the development of a Hybrid RF+GRU model that combines the feature extraction capability of RF with the temporal learning strength of GRU for enhanced forecasting performance. The proposed approach also integrates effective pre-processing techniques, feature extraction, and SMOTE-based data balancing to improve model reliability and prediction efficiency in pharmaceutical sales analysis.

1.2. Organization of the Paper

The structure of the paper is as follows: Section II reviews related work on Pharmaceutical Sales Analysis and Forecasting, Section III details the dataset, pre-processing procedures, and model implementation, Section IV provides the experimental results along with comparative analysis, and Section V concludes the study by highlighting key findings and suggesting directions for future research.

2. Literature Review

This section reviews and analysis of the major research in the field of Pharmaceutical Sales Analysis and Forecasting were done in order to bring in and enrich the analysis of this study.

Pang et al. (2025) The aim of the study proved that the best model for making predictions rests on the features of the service demand data. The N02BE and R06 classes were best served by the SARIMA model (RMSE: 57.46 and 9.24, respectively), but the M01AE class was better served by the Prophet model (RMSE: 9.52). With the largest volatility (Std. Dev: 76.07) and, thus, the biggest inventory risk, Change Point Detection found a notable structural break in demand for N02BE [23].

Dinh, Do and Nguyen, (2024) proposed forecasting approach outperforms the traditional ARIMA model in pharmaceutical demand prediction. The ARIMA model achieved an MAE of 7.8% and an RMSE of 10.1%, whereas the LR model showed improved forecasting ACC with a lower MAE of 6.4% and RMSE of 8.7%. These findings indicate that AI-driven predictive techniques can provide more reliable pharmaceutical forecasting performance, supporting better inventory planning, market analysis, and healthcare decision-making[24].

Qassrawi, Azzeh and Hijjawi, (2024) Experimental results showed that Long Short-Term Memory outperformed MLP and CNN in producing sales predictions with an average RMSE of 1.28(k) and a Mean Absolute Error of approximately 0.85(k), as well as in predicting USD prices with an average Root Mean Square Error of approximately 0.75 and a MAE of about 0.44. The projections are subsequently utilized to modify inventory levels in accordance with the forecasts [25].

Dutta, Das and Chatterjee, (2022) attempt to evaluate five distinct ML algorithms on the pharmaceutical product dataset and settle on linear regression as the most effective. A superior MAPE of 19.07% indicates that it outperforms competing models. Research confirms that linear regression is the most accurate model for forecasting sales of medicinal products [26].

Konar and Pitroda, (2022) obtained for guessing how likely it is that someone use internet drugs after The following are the COVID results: an ACC of around 85% was achieved by the gradient booster model, an ACC of around 71% by the DT classifier, an ACC of about 85% by the LR model, and an ACC of about 78% by the RF model. Decision tree classifiers achieved an ACC of around 78%, logistic regression about 50%, and support vector machines about 64% when it comes to predicting the likelihood that a person utilizes at-home lab testing after COVID [27].

Mbonyinshuti et al., (2021) centered on using ML to predict patterns in Rwanda's crucial medicine demand going forward. An artificial neural network, an RF, and a linear regression model were developed and used. The RF allows users to input a month, year, district, and medicine name in order to predict future demand using consumption statistics. With the train set, it was able to predict 10 different medications with an ACC of 88%, and with the test set, it was 76% accurate [28].

Table I provides an overview of recent studies on Pharmaceutical Sales Analysis and Forecasting, including the proposed models, key results, and challenges encountered.

Table 1: Recent Studies on Pharmaceutical Sales Analysis and Forecasting using Machine Learning technique

Author	Proposed Work	Results	Key Findings	Limitations & Future Work
Pang et al. (2025)	Used SARIMA and Prophet models for pharma demand forecasting.	SARIMA gave best RMSE for N02BE (57.46) and R06 (9.24); Prophet best for M01AE (9.52).	Forecasting accuracy depends on demand pattern and volatility.	Future work can use hybrid AI models and larger datasets.
Dinh, Do and Nguyen (2024)	Compared ARIMA and Linear Regression for pharmaceutical forecasting.	LR achieved MAE 6.4% and RMSE 8.7%, better than ARIMA.	AI models improve forecasting accuracy and inventory planning.	Future studies can apply deep learning methods.
Qassrawi, Azzeh and Hijjawi (2024)	Applied LSTM, CNN, and MLP for sales and price prediction.	LSTM achieved lowest RMSE and MAE values.	LSTM performed best for sequential pharma sales prediction.	Future work can include economic and seasonal factors.
Dutta, Das and Chatterjee (2022)	Tested five ML algorithms for pharma product sales prediction.	Linear Regression achieved best MAPE of 19.07%.	Linear Regression was most effective among tested models.	Advanced ensemble and deep learning models can be explored.
Konar and Pitroda (2022)	Predicted online pharmacy and lab test usage post-COVID using ML.	Gradient Booster and Logistic Regression achieved ~85% accuracy.	ML models effectively predicted healthcare consumer behavior.	Future work can use larger datasets and more features.
Mbonyinshuti et al. (2021)	Used ML models for drug demand forecasting in Rwanda.	RF achieved 88% train accuracy and 76% test accuracy.	Random Forest showed effective prediction performance.	Limited dataset; future work can use advanced hybrid models.

Research gaps: Traditional ML and statistical models have dominated the field of pharmaceutical sales forecasting thus far, but these approaches fail to account for the intricate sequential and nonlinear correlations seen in pharmaceutical demand data. Previous studies also had small sample sizes, limited predictive power, and failed to account for data imbalance and feature optimization. Besides, most studies have employed a single forecasting model and have not fully integrated the feature extraction and temporal learning features. Thus, an efficient hybrid forecasting approach that can enhance the ACC of the forecasts, minimize the forecasting errors, and offer a reliable analysis of pharmaceutical sales using advanced ML and DL techniques is necessary.

3. Research Methodology

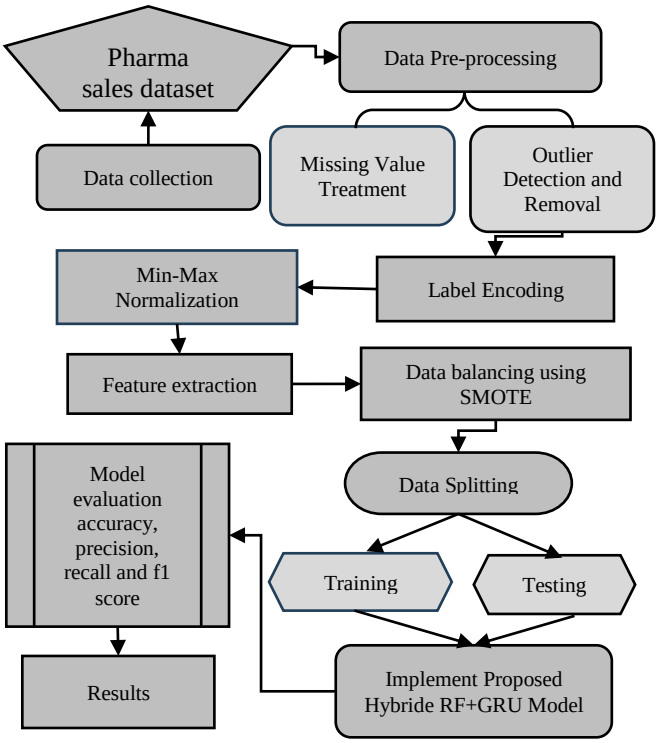

Fig 1: Proposed Flowchart for Pharmaceutical Sales Analysis and Forecasting Using Machine Learning

In this study, a hybrid model of RF and GRU is developed and applied to a Pharma sales dataset obtained from Kaggle for the pharmaceutical sales analysis and forecasting. To enhance the data quality and prediction performance, data

preprocessing techniques like missing value processing, outlier detection, data normalization, feature extraction and SMOTE data balancing were used. The proposed model is an amalgam of RF and GRU, which enables it to acquire characteristics and employ sequential forecasting for accurate pharmaceutical sales prediction. The model's performance is assessed using metrics including F1, REC, ACC, and PRE. The proposed ML-based system for analyzing and forecasting pharmaceutical sales is depicted in Figure 1 as a flow diagram.

A comprehensive breakdown of the suggested procedure is provided in the section that follows.:

3.1. Data Gathering and Analysis

The Pharma sales dataset is used in this investigation. The information was culled from 57 different pharmaceutical items' weekly sales records spanning 2014–2019 and included 600,000 entries. The brand name, quantity sold, and date of sale were all part of the record. The following are examples of data visualizations that were employed to analyze the distribution of attacks, feature correlations, etc.: bar plots and heatmaps:

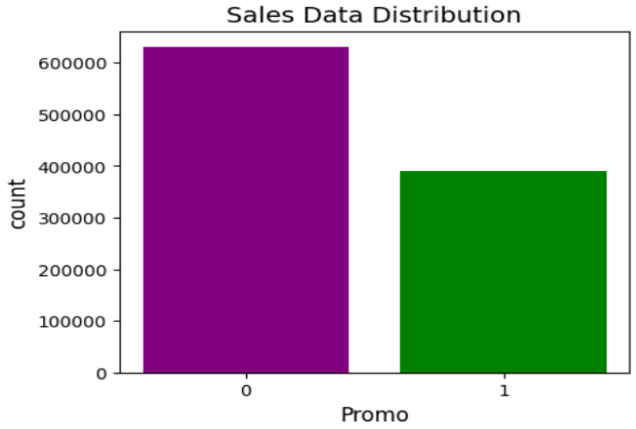

Fig 2: Bar graph of class distribution of Pharma sales dataset

The bar graph in Figure 2 shows the distribution of classes in the Pharma sales dataset, broken down by promotional and non-promotional sales categories. As seen in the graph, there are many more non-promotional sales instances than there are promotional sales instances, which suggests that there is not a lot of balance in the data set. This visualization is useful to grasp the general sales structure, and how the majority of the data consists of regular sales records.

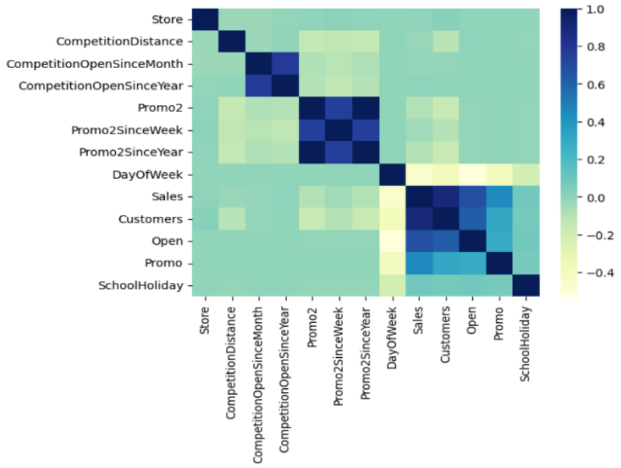

Fig 3: Correlation Matrix Heatmap on Pharma Sales Dataset for Pharmaceutical Sales Analysis

Figure 3 displays the Pharma sales dataset's correlation matrix heatmap, which demonstrates the correlations among characteristics. A stronger positive correlation is indicated by a brighter color, and a stronger negative correlation by a darker one. The heatmap displays that Customers, Open and Promo are positively associated with Sales and are important features for pharmaceutical sales forecasting.

3.2. Data Pre-processing

The Pharma sales data set was used to perform data preparation tasks such as data integration, cleaning, and feature extraction. Various data preprocessing steps were performed including handling missing data, outlier detection and removal, data labeling and normalization to enhance data quality and consistency. The following are the key steps in preprocessing performed in this study:

- Missing Value Treatment: The data quality and reliability of the model were improved by handling null values in the Pharma sales dataset through the missing value treatment process.
- Outlier Detection and Removal: To detect and remove any outliers or inconsistencies in the Pharma sales data, it underwent outlier detection and removal. Outlier detection and removal were conducted to make sure that there are no outliers or inconsistencies in the data that would impact the ACC and stability of the prediction model.
- Label Encoding: Categorical attributes (e.g., sales categories, class labels) were encoded into numerical values, ensuring they could be processed by ML algorithms.

3.3. Min-Max Normalization

Min-max normalization is a type of feature scaling that rescales numerical data to a set interval, usually 0 to 1. This approach enables the variables to be analyzed without any one variable dominating the analysis because of the large number of values it has. It enhances the consistency, comparability, and overall efficiency of ML algorithms. The normalization was carried out based on the following mathematical formula (1):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \ldots\ldots\ldots (1)$$

Where X represents the feature value, X' stands for the feature's normalized value, X_{min} denotes the feature's smallest value, and X_{max} denotes its highest value.

3.4. Feature Extraction

Enhancing the performance of ML models may be achieved by feature extraction, which entails choosing and modifying data in order to transform it into features. In order to successfully predict future medication sales, feature extraction is used to identify and remove the most relevant characteristics from pharmaceutical sales data. This enables the original data to be converted to a more useful, smaller representation of data, while maintaining all the information necessary to make good forecasts. Feature extraction improves the efficiency and performance of ML models by removing irrelevant features and simplifying the data. This also helps in better understanding of vital sales patterns and trends, which leads to better drug sales predictions.

3.5. Data balancing using Synthetic Minority Over-sampling Technique (SMOTE)

A technique for manipulating imbalanced datasets in which some classes occur much more frequently than others, so improving the performance of ML models is called data balancing. In this research, synthetic minority class oversampling (SMOTE) method was used to create synthetic minority samples and balance the distribution of classes. SMOTE generates synthetic examples by determining the Euclidean distance between minority class examples and their most similar examples.

3.6. Data Splitting

The data set was split into a training set and testing set in a proportion of 90:10 with stratification. This was done to preserve the proportional distribution of classes in both sets, similar to the original distribution of the data set.

3.7. Proposed Hybrid Random Forest and Gated Recurrent Unit (RF+GRU) Model

The proposed Hybrid RF+GRU Model is designed for the Pharmaceutical Sales Analysis and Forecasting for this work. The proposed model is a Hybrid RF+GRU model that blends the strengths of both RF and GRU models, enhancing the ACC of pharmaceutical healthcare forecasting. The RF model is applied to extract important features based on the decisions and the GRU model is applied to learn temporal and sequential dependencies from the data set. The combination of these two models has better forecasting ACC, lower prediction error and model stability.

3.7.1. Random Forest (RF) Model

RF is a kind of ensemble learning that takes the output of several decision trees and utilizes them to generate a single forecast. By enhancing generalizability and decreasing overfitting, the RF model effortlessly manages complicated healthcare datasets. The methodology for training the RF model with pharmaceutical healthcare data is illustrated in Equation (1). The final forecast is obtained by averaging the outputs of all DTs.

$$\hat{Y}_{RF} = \frac{1}{N}\sum_{i=1}^{N} T_i(X) \ldots\ldots\ldots (1)$$

In this context, $\hat{Y}_{RF}$ signifies the RF model's ultimate forecast, N stands for the total count of trees, and $T_i(X)$ shows the forecast produced by the i^{th} Given input data X. By reducing variation and increasing forecast stability, the averaging process is a useful tool.

3.7.2. Gated Recurrent Unit (GRU) model

A RNN architecture, the GRU model is built to detect patterns in data that occur sequentially and over time. To avoid vanishing gradient issues and control the flow of information, GRU employs gating methods. Accurate forecasting is achieved in the proposed study by training the GRU model with pharmaceutical healthcare records, which reveal long-term dependencies. Equation (2) shows that the final output of the network is generated by integrating both past and current information, with the update gate controlling the flow of past information and the reset gate regulating the forgetting mechanism. This formulation collectively represents the three components of the Gated Recurrent Unit (GRU).

$$h_t = (1 - z_t)\odot h_{t-1} z_t + \tanh\odot(W[h_{t-1}, x_t]) \ldots\ldots\ldots (2)$$

The hidden state at time t is represented by h_t, The input vector at time t is x_t, the update gate regulating the amount of past information is z_t, and the weight matrix associated with the model is W. In order to make the network non-linear, the activation function is tanh. The GRU architecture effectively captures sequential dependencies in healthcare data and improves the ACC of pharmaceutical sales forecasting.

3.7.3. Hybrid RF+GRU Model

The Hybrid RF+GRU model combines the feature extraction ability of RF and the temporal learning ability of GRU. At first, the RF model detects important features of the healthcare data and makes initial predictions. These optimized outputs are then fed into the GRU which is used for sequential learning and final forecasting. The hybrid model's final prediction is calculated as shown in Equation (3).

$$Y_{Hybrid} = \alpha\hat{Y}_{RF} + \beta\hat{Y}_{GRU} \ldots\ldots\ldots (3)$$

Y_{Hybrid} is the final prediction by the hybrid model, while $\hat{Y}_{RF}$ and $\hat{Y}_{GRU}$ are the predictions of the RF model and GRU model, respectively, and α and β are the weighting coefficients to balance the contributions of the two models. The combination of the two methods ensures better prediction ACC, effectiveness, and overall better performance of healthcare data analysis.

Optimized hyperparameters were used to implement the proposed Hybrid RF+GRU model to enhance forecasting performance. RF model was given 100 estimators, max depth of 20 and min sample split of 2. To achieve efficient pharmaceutical sales prediction, the learning rate used for the GRU model was set at 0.001, with a batch size of 64, number of epochs of 100, and hidden units of 128.

3.8. Evaluation Metrics

The suggested approach was evaluated using a number of widely used performance metrics. First, developed a confusion matrix to display the classification results, which include the total number of occurrences of each class that were properly and wrongly categorized. Extracted TP, FP, TN, and FN, which are the main elements of this array. Following the explanation of these values, important assessment metrics such as ACC, PRE, REC, and F1 were computed:

Accuracy: This metric measures how well the trained model predicted a collection of instances relative to the total number of input samples in the dataset. Equation (4) gives us this information-

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{Fp} + \text{TN} + \text{FN}} \ldots\ldots (4)$$

Precision: One way to evaluate a model's performance is to divide the total number of positive occurrences by the number of true positives that the model correctly predicts. This ratio is called PRE. Equation (5) expresses it, and it shows how well the classifier finds positive classifications-

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \ldots\ldots (5)$$

Recall: The ratio of genuine positives to the sum of true positives and true negatives was calculated. This metric, which can be expressed mathematically as Equation (6), assesses the model's capacity to identify genuine positive occurrences-

$$Recall = \frac{\text{TP}}{TP + FN} \ldots\ldots\ldots (6)$$

F1 score: This is the harmonic mean of PRE and REC, giving a balanced score of the two. The value of this metric is between 0 and 1. It is mathematically expressed as Equation (7)-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \ldots\ldots\ldots (7)$$

4. Results and Discussion

The experimental setup and evaluation of the proposed model are detailed in this section. The model is tested using pharmaceutical sales data in particular. The tests were conducted on a state-of-the-art HPC system that utilized a Linux operating system, an Intel Core 2 Duo processor with a 2.16 GHz clock speed, and 8 GB of RAM for system design and analysis. Table II shows the classification performance of proposed Hybrid RF+GRU model for Pharmaceutical Sales Analysis and Forecasting on the Pharma Sales dataset. The outcomes demonstrate the strong predictive capabilities of the model when evaluated according to all the evaluation metrics used, with an ACC of 96.1% indicating the high overall correctness, PRE of 96.8% indicating the low false positive rate, REC of 96% indicating the effectiveness of the identification of actual positive instances, and an F1 of 96.4% indicating a proper trade-off between PRE and REC. Overall, these findings confirm that the proposed hybrid RF+GRU model has strong and trustworthy forecasting capabilities in the pharmaceutical sector.

Table 2: Classification Results of Proposed Hybrid RF+Gru Model Pharmaceutical Sales Analysis and Forecasting Using Pharma Sales Dataset

Matrix	Proposed Hybrid RF+GRU Model
Accuracy	96.1
Precision	96.8
Recall	96
F1-score	96.4

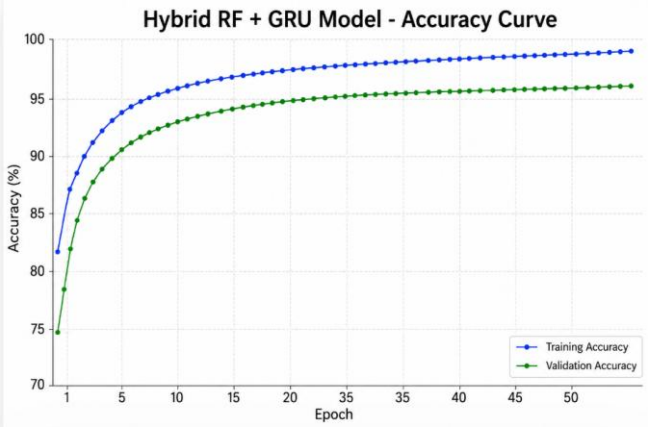

Fig 4: Accuracy Curve for Proposed Hybride RF+GRU Model

Training and validation ACC improve as the number of epochs increases, as seen in the ACC curve of the Hybrid RF+GRU model in Figure 4. Following a fast increase in the first few epochs, the model's ACC begins to converge more slowly, which means that it has learned and stabilized. Also, the validation ACC is consistently rising, which indicates effective generalization and overfitting mitigation as well as a slight departure from the training curve.

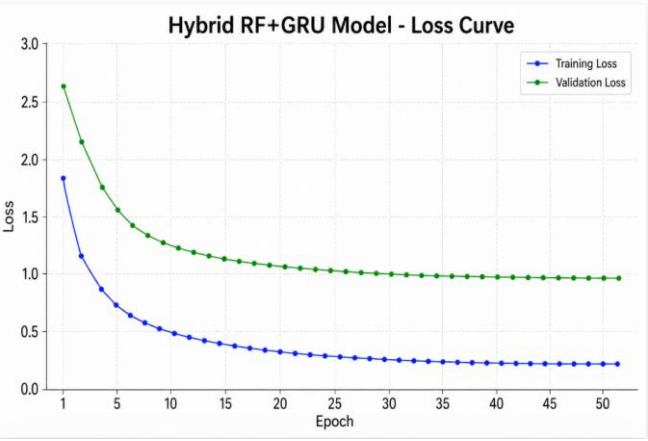

Fig 5: Loss Curve for Proposed Hybride RF+GRU Model

The loss curve of the Hybrid RF+GRU model shows that the training and validation losses decrease gradually across the epochs, as shown in Figure 5. This suggests that the model is learning and converging well. Excellent fitting performance on the training data is demonstrated by the training loss, which rapidly declines and remains below the validation loss. Good generalizability and reduced prediction error are shown by the stabilization of the validation loss after a few epochs.

4.1. Comparative Analysis

Table III presents a comparative ACC evaluation using current models to evaluate the efficacy of the proposed Hybrid RF+GRU model. Detailed comparisons of many ML models for pharmaceutical sales analysis and forecasting are shown in Table III, using the pharma sales data set. Compared to all of the baseline models, the suggested Hybrid RF+GRU model significantly outperforms them in every measure used for evaluation. To be more specific, the SVM model boasts an F1 of 84%, an ACC of 82%, a PRE of 84%, and a REC of 79%, while the XGBoost model manages 68.8% ACC, 81.6% PRE, and 69.5% REC. Despite not providing complete metric values, the KNN model achieves an ACC of 86.6%. In contrast, the suggested Hybrid RF+GRU model is remarkably balanced across all assessment metrics, with a 96.1% ACC, 96.8% PRE, 96% REC, and 96.4% F1. Finally, when compared to traditional ML models, the suggested hybrid solution outperforms them in terms of ACC and reliability, proving its efficacy in pharmaceutical sales forecasting.

Fig 6: Comparison of Different Machine Learning Models for Pharmaceutical Sales Analysis and Forecasting

Model	Accuracy	Precision	Recall	F1-score
XGBoost[29]	68.8	81.6	69.5	-
SVM[30]	82	84	79	84
KNN[31]	86.6	-	-	-
Proposed Hybrid RF+GRU Model	96.1	96.8	96	96.4

The model proposed was the Hybrid RF+GRU with an ACC of 96.1% in evaluating and predicting the product sales in pharmaceuticals. It performs effective feature selection by RF and sequential learning by time-dependency using a GRU. This combination is better than using each model separately for making predictions and avoiding overfitting. Reliability, stability, and efficiency in forecasting are all improved by the suggested approach, according to the results.

5. Conclusion and Future Study

Predicting future pharmaceutical demand is essential for healthcare systems to save money, reduce waste, and make sure drugs are available when needed. In recent years, ML technologies have been incredibly helpful in tackling the difficult and extremely nonlinear demand forecast in the pharmaceutical sector. This research proposes an efficient Pharmaceutical Sales Analysis and Forecasting system based on a Hybrid RF+GRU model, developed on the Pharma sales dataset available on Kaggle. Pre-processing methods such as data balancing and feature extraction are used to enhance the model performance and forecasting ACC. The experimental outcomes indicated that the best ACC of 96.1% was obtained by the proposed RF+GRU model compared to the other models, XGBoost, SVM, and KNN, which are able to capture the important features of sales and sequential patterns. In conclusion, the proposed hybrid model is a promising and effective approach to addressing the challenges of forecasting pharmaceutical sales and decision-making in healthcare settings. Future work could involve extending the model with larger real-time healthcare datasets and using more sophisticated DL optimization methods to further enhance forecasting ACC.

References

[1] S. Dharmavaram and P. Bhanushali, "Machine Intelligence-Driven Forecasting for ED Triage and Dynamic Hospital Patient Routing," Feb. 2026. doi: 10.64898/2026.02.18.26346566.

[2] A. G. Kravets, M. A. Al-Gunaid, V. I. Loshmanov, S. S. Rasulov, and L. B. Lempert, "Model of medicines sales forecasting taking into account factors of influence," *J. Phys. Conf. Ser.*, vol. 1015, p. 032073, May 2018, doi: 10.1088/1742-6596/1015/3/032073.

[3] M. R. Anand, "Enhancing Pharmaceutical Supply Chains with Densenet-121 and 1D CNN Integration," in *2025 Global Conference in Emerging Technology (GINOTECH)*, IEEE, May 2025, pp. 1–7. doi: 10.1109/GINOTECH63460.2025.11076754.

[4] J. Jiménez-Luna, F. Grisoni, N. Weskamp, and G. Schneider, "Artificial intelligence in drug discovery: recent advances and future perspectives," *Expert Opin. Drug Discov.*, vol. 16, no. 9, pp. 949–959, Sep. 2021, doi: 10.1080/17460441.2021.1909567.

[5] R. S. Snehamruth, "Data-Driven Optimization of Pharmaceutical Manufacturing Processes using Quality by Design (QbD) Frameworks," *Int. J. Curr. Eng. Technol.*, vol. 14, no. 6, pp. 557–566, 2024, doi: 10.14741/ijcet/v.14.6.19.

[6] P. Kumar, "Edge Computing and IoT for Real-Time Healthcare Data Processing and Integration," in *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, IEEE, Dec. 2025, pp. 105–110. doi: 10.1109/ICAAIC64647.2025.11331211.

[7] C. Tayal and S. Murumkar, "Patient Identity Protection and Duplicate Record Prevention in Electronic Health Record (EHR) Systems," in *2026 18th International Conference on Knowledge and Smart Technology (KST)*, 2026, pp. 458–464. doi: 10.1109/KST67832.2026.11431915.

[8] S. Besma, C. Rachid, and K. Abdelaziz, "For an Effective Management of the Functional Capacities of Companies: A Study of Pharmaceutical Companies," *Int. J. Saf. Secur. Eng.*, vol. 11, no. 5, pp. 557–563, Oct. 2021, doi: 10.18280/ijsse.110507.

[9] F. Silva-Aravena, I. Ceballos-Fuentealba, and E. Álvarez-Miranda, "Inventory Management at a Chilean Hospital Pharmacy: Case Study of a Dynamic Decision-Aid Tool," *Mathematics*, vol. 8, no. 11, p. 1962, Nov. 2020, doi: 10.3390/math8111962.

[10] J. W. Sajja and G. B. Komarina, "Enhancing compliance and data integrity in life sciences and healthcare with S/4HANA: A data management and governance framework," *World J. Adv. Eng. Technol. Sci.*, vol. 15, no. 2, pp. 2816–2827, May 2025, doi: 10.30574/wjaets.2025.15.2.0843.

[11] S. Mahmud, "AI and Data Analytics for Enhancing Home

Healthcare: Optimizing Patient Outcomes and Resource Allocation," *Front. Appl. Eng. Technol.*, vol. 2, no. 1, pp. 23–100, 2025, doi: 10.70937/faet.v2i01.61.

[12] J. A. Kachhia, "Healthcare Predictive Analytics Based on Machine Learning Techniques for Identifying Cardiovascular Risk Screening," *Int. J. Curr. Eng. Technol.*, vol. 13, no. 6, pp. 635–642, Dec, 2023, doi: https://doi.org/10.14741/IJCET/V.13.6.17.

[13] S. Golriz Khatami, S. Mubeen, V. S. Bharadhwaj, A. T. Kodamullil, M. Hofmann-Apitius, and D. Domingo-Fernández, "Using predictive machine learning models for drug response simulation by calibrating patient-specific pathway signatures," *npj Syst. Biol. Appl.*, vol. 7, no. 1, p. 40, oct. 2021, doi: 10.1038/s41540-021-00199-1.

[14] K.-K. Mak and M. R. Pichika, "Artificial intelligence in drug development: present status and future prospects," *Drug Discov. Today*, vol. 24, no. 3, pp. 773–780, Mar. 2019, doi: 10.1016/j.drudis.2018.11.014.

[15] P. Kelle, J. Woosley, and H. Schneider, "Pharmaceutical supply chain specifics and inventory solutions for a hospital case," *Oper. Res. Heal. Care*, vol. 1, no. 2–3, pp. 54–63, Jun. 2012, doi: 10.1016/j.orhc.2012.07.001.

[16] A. Aliper *et al.*, "Prediction of Clinical Trials Outcomes Based on Target Choice and Clinical Trial Design with Multi-Modal Artificial Intelligence," *Clin. Pharmacol. Ther.*, vol. 114, no. 5, pp. 972–980, 2023, doi: 10.1002/cpt. 3008.

[17] P. Kumar, "Leveraging Generative AI for Automated Data Standardization and Interoperability in Healthcare," in *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, IEEE, Dec. 2025, pp. 99–104. doi: 10.1109/ICAAIC64647.2025.11330217.

[18] M. R. Anand and A. K. S, "Temporal Fusion Transformer Forecasting and MILP Prescriptive Optimization for Hospital Pharmacy Supply Chain Orchestration," in *2025 9th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Nov. 2025, pp. 1206–1213. doi: 10.1109/ICECA66444.2025.11382695.

[19] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, no. July 2019, p. 106435, May 2020, doi: 10.1016/j.cie.2020.106435.

[20] M. Azadi, S. Yousefi, R. Farzipoor Saen, H. Shabanpour, and F. Jabeen, "Forecasting sustainability of healthcare supply chains using deep learning and network data envelopment analysis," *J. Bus. Res.*, vol. 154, p. 113357, Jan. 2023, doi: 10.1016/j.jbusres.2022.113357.

[21] R. Snehamrutha, "Patient Engagement Strategies in Community Pharmacies and their Effect on Vaccination Uptake and Medication Synchronizations," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 3, pp. 163–173, 2023, doi: 10.56472/25832646/JETA-V3I7P120.

[22] M. R. Anand and K. Abhilash, "Transforming Energy-Intensive Smart Factories with AI: TCN-based Forecasting and DQN-Driven Operational Optimization for Healthcare Manufacturing," in *International Conference on Intelligent Computing, Information and Control Systems (ICOIICS-2025)*, IEEE, 2025, pp. 508–515, November. doi: 10.1109/ICOIICS67115.2025.11390244.

[23] F. Pang, X. Zhou, T. Bai, K. Wen, J. Zhu, and B. Wu, "A multi-level time-series analysis for forecasting and operational planning in pharmaceutical services: A case study," Jul. 2025. doi: 10.21203/rs. 3. rs-7114272/v1.

[24] H. N. Dinh, T. H. Do, and T. B. Nguyen, "An Efficiency Improvement of the N-Beats Model for Sale Forecast Problem," in *Creative Approaches Towards Development of Computing and Multidisciplinary IT Solutions for Society*, Wiley, 2024, pp. 251–263. doi: 10.1002/9781394272303.ch15.

[25] N. Qassrawi, M. Azzeh, and M. Hijjawi, "Drug sales forecasting in the pharmaceutical market using deep neural network algorithms," *Int. J. Syst. Innov.*, vol. 8, no. 3, pp. 63–83, 2024, doi: 10.6977/IJoSI.202409_8(3).0006.

[26] S. R. Dutta, S. Das, and P. Chatterjee, "Smart Sales Prediction of Pharmaceutical Products," in *2022 8th International Conference on Smart Structures and Systems (ICSSS)*, IEEE, Apr. 2022, pp. 1–6. doi: 10.1109/ICSSS54381.2022.9782271.

[27] K. Konar and H. Pitroda, "Analyzing and Predicting the Impact of COVID-19 on Online Pharmaceuticals Sectors and Pathological Services in India," in *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2022, pp. 1–7. doi: 10.1109/I2CT54291.2022.9824883.

[28] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, "The Prediction of Essential Medicines Demand: A Machine Learning Approach Using Consumption Data in Rwanda," *Processes*, vol. 10, no. 1, p. 26, Dec. 2021, doi: 10.3390/pr10010026.

[29] R. Pall, Y. Gauthier, S. Auer, and W. Mowaswes, "Predicting drug shortages using pharmacy data and machine learning," *Health Care Manag. Sci.*, vol. 26, no. 3, pp. 395–411, set. 2023, doi: 10.1007/s10729-022-09627-y.

[30] Y. Xiong, "Development of an AI-Driven Model for Drug Sales Prediction Using Enhanced Golden Eagle Optimization and XGBoost Algorithm," *Informatica*, vol. 49, no. 17, Mar. 2025, doi: 10.31449/inf.v49i17.7491.

[31] M. Husban, A. Mir, and I. Yustiana, "Predicting Big Mart Sales with Machine Learning," in *The 7th International Global Conference Series on ICT Integration in Technical Education & Smart Society*, Basel Switzerland: MDPI, Sep. 2025, p. 95. doi: 10.3390/engproc2025107095.