Journal of Global Research in Multidisciplinary Studies (JGRMS)
Volume 2 | Issue 7 | July 2026 | Pages 29-36
DOI: https://doi.org/10.5281/zenodo.21836894
e-ISSN: 3049-2963
RESEARCH ARTICLE OPEN ACCESS
Available online: https://saanvipublications.com/journals/index.php/jgrms

Artificial Intelligence-Assisted Pharmaceutical Manufacturing Optimization Using Predictive Machine Learning Techniques

Dr Chintal Kumar Patel
Associate Professor, CSE
Geetanjali Institute of Technical Studies
chintal.patel@gits.ac.in
Corresponding Author*Dr.Chintal Kumar Patel (chintal.patel@gits.ac.in)

Received: 18 Jujy 2026 | **Revised:** 22 July 2026 | **Accepted:** 25 July 2026 |

***Abstract*—The pharmaceutical manufacturing sector is seeing growing difficulties in streamlining operations, preserving product quality, and meeting stringent regulatory standards. Now more than ever, AI and ML are poised to revolutionize pharmaceutical production by evaluating data and generating predictions to optimize various elements. Utilizing a combination of AI-driven approaches, massive volumes of industrial data may be employed for process monitoring, quality prediction, defect detection, and production optimization. Using machine learning techniques like the XGBoost algorithm in conjunction with the SECOM production Process Dataset from UCI, this research suggests a smart framework for optimizing pharmaceutical production. Here are some potential approaches for data cleaning: Data balancing, feature scaling, dimensionality reduction, and missing data imputation are the four pillars of ADASYN, a methodology that improves the precision and effectiveness of data models. Experimental results show that the suggested XGBoost model achieves better results than competing models in terms of ACC (98.93%), PRE (98.97%), REC (98.99%), and F1 score (98.92%). These models include SVM, RF, DT, DNN, and MLP. Improvements in quality control, process efficiency, and intelligent decision-making are achieved via the use of the AI-based framework, which is proven to be a reliable way for enhancing predictive pharmaceutical manufacturing.**

***Keywords*—Artificial Intelligence, Machine Learning, Pharmaceutical Formulation, Drug Delivery Systems, Quality by Design, Pharmaceutical Manufacturing.**

I. INTRODUCTION

The healthcare and pharmaceutical sectors are crucial to better world health because they guarantee the creation, production, and distribution of medications that are safe, effective, and of high quality [1][2][3]. The manufacturing of the pharmaceuticals is complex in nature and involves formulation design, selection of raw material, process optimization, quality assurance, and regulatory compliance [4][5][6]. The basic approach of conventional manufacturing and formulation development is based on a lot of experiments in the lab and subsequent validation tests to establish optimal processing conditions [7]. Product development cycles and production costs might be lengthened and increased when using traditional methods because of their time, energy, and money requirements [8][9].

In this context, AI has become an enabling technology that can help overcome these challenges by leveraging data-driven decision making in pharmaceutical research and manufacturing [10][11]. AI analyzes massive amounts of data in the industrial and pharmaceutical industries by combining computer intelligence, pattern recognition, and predictive analytics [12]. Automated systems with AI can help in quality assurance, process control, stability prediction, and intelligent manufacturing, enhancing product uniformity and reliability [13]. Moreover, with the rise of pharmaceutical manufacturing data, sensor technologies, and industrial automation, the use of AI-powered manufacturing optimization has also been gaining momentum [14][15][16].

The use of DL and ML has been on the rise in recent years, which has improved the capacity for AI-driven optimization of pharmaceutical production processes through the development of very precise prediction models [17][18]. This is achieved by analyzing large-scale pharmaceutical manufacturing data using ML techniques to uncover patterns and enhance prediction ACC, and by automatically extracting complex features from the data, thereby optimizing formulations, assessing quality, and monitoring production processes in real time with DL models [19]. The incorporation of ML and DL increases production efficiency, decreases the time needed for development and manufacturing, and increases the quality and compliance of products [20][21].

A. Motivation and Contribution

The pharmaceutical industry's need for complicated but reliable, efficient, and cost-effective manufacturing systems is the driving force behind this study. Traditional manufacturing approaches require extensive experimentation, consume significant resources, and often face challenges in achieving consistent quality and optimized performance. Advanced analytics and capabilities of large-scale manufacturing data, computer simulation of process results and the ability to make

Copyright © 2026 The Author(s).
This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, adaptation, and reproduction in any medium, provided the original author(s) and the source are properly credited.

J. Glob. Res. Multidiscip. Studies, 2 (7):29-36
DOI: https://doi.org/10.5281/zenodo.21836894

more accurate decisions are made possible by AI and ML-based technologies. The study's goal is to create an intelligent optimization framework based on XGBoost that improves the efficiency, quality control, and reliability of pharmaceutical manufacturing. The following are a few of the most important contributions made by this study:

- Developed an AI-driven intelligent pharmaceutical manufacturing optimization framework using the proposed XGBoost ML model for accurate process prediction and quality improvement.
- Utilized the UCI SECOM Manufacturing Process Dataset containing sensor-based manufacturing measurements to analyze and optimize pharmaceutical production processes.
- Enhanced data quality and model performance through the implementation of effective preprocessing approaches. These techniques include managing missing data, dimensionality reduction, feature encoding, normalization, and ADASYN-based data balancing.
- The model's ability to detect failure scenarios was enhanced by creating synthetic minority samples to rectify the class imbalance issue in manufacturing data.
- Preformed a comparison study of the existing models (SVM, RF, DT, DNN and MLP) and demonstrated that the proposed technique is superior.
- Provided a reliable AI-powered decision support tool for intelligent pharmaceutical manufacturing optimization, improving quality control, process monitoring and operational efficiency.

The significance of the proposed research lies in the need for an intelligent and data-driven framework that can help reduce production failures, enhance quality control, and improve pharmaceutical manufacturing processes. The novelty of this work lies in the ability to integrating advanced preprocessing techniques with the XGBoost algorithm to solve high dimensional sensor data and class imbalance problems. The proposed method has more prediction capability than existing ML and DL models with the accurate manufacturing outcome identification. The system offers reliable, effective immediate optimization and decision making in pharmaceutical processes.

B. Organization of the Paper

The paper is structured as follows: A review of relevant literature is given in Section II. The suggested method's architecture is presented in Section III. In Section IV, detail the experimental setup and go over the assessment outcomes. The article is wrapped up in Section V, which discusses the study limits and future directions.

II. LITERATURE REVIEW

An extensive survey of the literature on intelligent pharmaceutical manufacturing optimization was carried out to identify recent advancements, research gaps, and opportunities for developing the proposed framework.

Thupphae and Wiwatwattana (2026) Proposed ML approaches were investigated to predict dissolution outcomes of coated tablets using manufacturing data from 85 production batches. Using layered cross-validation, tested many regression models: support vector, linear, ridge, lasso, elastic net, and random forest. With $R^2 = 0.419$, MAE = 1.70, and RMSE = 2.27, support vector regression outperformed the other methods. Permutation feature importance and SHAP analysis identified the particle size distribution of croscarmellose sodium as the most influential factor [22].

Mannam (2026) Introduces a data-driven strategy for anticipating OOS hazards through the use of LIMS, audit trails from the past, and environmental data. An AUC of 0.91 was attained using a supervised learning model based on XGBoost that made use of attributes that were particular to time, instruments, and analysts. SHAP values enhance explainability by identifying key feature contributions, enabling proactive risk assessment and supporting GMP-compliant quality control [23].

M et al. (2025) DL and AI are combined in an AI-driven smart control system that optimizes the brazing process, detects faults, and enables predictive maintenance. This system targets these difficulties. With the suggested system, and cut energy consumption by 18.7%, increase process efficiency by 23.5 percent, and obtain ACC for fault detection of 97.2 percent. The processing overhead is a constraint, though, therefore there has to be future work done to make it more scalable and real-time adaptable for industrial application [24].

Bhende et al., (2025) offers a smart control system that uses AI and ML/DL to improve brazing quality, find flaws, and provide predictive maintenance. Convolutional neural networks (CNNs) are good at defect classification, but they can also spot outliers for real-time deviations. With this technology, and able to cut energy usage by 18.7%, improve the efficiency of the defect detection process by 23.5%, and identify fault rates up to 97.2%. A larger industrial acceptance might be possible if future research focused on improving the scalability and real-time flexibility, as computational overhead is a restriction [25].

Kakavandi et al.(2025) synthetic sample creation using a simulation model and deep generative models. The quantitative and qualitative assessment of simulations based on FEA and VAE. The results indicate that different techniques perform differently in the sense that most of the synthetic data has reduced training iterations and improved prediction performance which leads to more accurate fault detection results. Interestingly, VAE synthesized normal samples gain the F1 of 0.185 for the fault detection model [26].

Malathi et al.(2024) explores the problems with the present pharmaceutical quality assurance procedures, including limited real-time monitoring, insufficient data analysis, and an absence of analytics and prediction. The suggested big data strategy makes it easier to analyze data in real-time and run predictive models. Experimental findings show a reduction of 35% in waste, a decrease of 40% in downtime, and a detection rate of 95% for faults. It can handle complex data volumes, ensuring proactive quality controls and industry compliance with data privacy and security protocols [27].

Lam et al.(2023) To predict the need for essential medicines in Nghe An province, Vietnam, they applied four ML models: RF, LGBM, Histogram-Based Gradient Boosting, and XGBoost. RF outperformed the other models in their regression analysis of pharmaceutical transaction data, with an RMSE of 0.95 and R^2/Adjusted R^2 of 0.81. This discovery supports the concept of the use of RF in healthcare resource allocation and stock management [28].

J. Glob. Res. Multidiscip. Studies, 2 (7):29-36
DOI: https://doi.org/10.5281/zenodo.21836894

Table I summarizes the recent literature on intelligent pharmaceutical manufacturing optimization by comparing the proposed models, datasets used, principal outcomes, and research challenges reported in previous studies.

TABLE I. SUMMERY OF THE STUDY ON INTELLIGENT PHARMACEUTICAL MANUFACTURING OPTIMIZATION WITH ML

Author	Proposed Work	Dataset	Results	Limitations & Future Work
Thupphae and Wiwatwattana (2026)	Predicted tablet dissolution using multiple ML regression models with explainable AI.	Manufacturing data from 85 coated tablet batches	SVR achieved the best performance (R^2 = 0.419, MAE = 1.70, RMSE = 2.27); SHAP identified key influential features.	Limited dataset and moderate predictive accuracy; future work should incorporate larger datasets and advanced hybrid models.
Mannam (2026)	Developed an XGBoost-based framework for predicting Out-of-Specification (OOS) risks using LIMS and audit data.	LIMS records, audit trails, and environmental data	Achieved AUC = 0.91 with SHAP-based explainability for proactive quality control.	Requires broader industrial validation and adaptation to diverse pharmaceutical manufacturing environments.
M et al. (2025)	Proposed an AI-driven smart control system integrating ML and DL for brazing optimization and predictive maintenance.	Industrial brazing process data	97.2% defect detection accuracy, 23.5% efficiency improvement, and 18.7% energy reduction.	High computational overhead; future work should improve scalability and real-time deployment.
Bhende et al. (2025)	Integrated CNN-based defect detection and anomaly detection for intelligent process monitoring.	Industrial manufacturing process data	97.2% defect detection accuracy with improved operational efficiency.	Scalability and real-time adaptability remain key research challenges.
Kakavandi et al. (2025)	Evaluated VAE and FEA-based synthetic data generation for industrial fault detection.	Industrial fault detection dataset with synthetic samples	VAE-generated samples improved fault detection (F1-score +0.185).	Synthetic data quality and model generalization require further investigation.
Malathi et al. (2024)	Applied big data analytics for pharmaceutical quality monitoring and predictive quality control.	Pharmaceutical manufacturing big data	95% fault detection, 40% downtime reduction, and 35% waste reduction.	Future work should strengthen intelligent decision support and advanced predictive analytics.
Lam et al. (2023)	Used ML models to forecast pharmaceutical demand for inventory optimization.	Pharmaceutical transaction data from Nghe An province	Random Forest achieved RMSE = 0.95 and R^2 = 0.81.	Limited regional applicability; future studies should validate models across diverse healthcare systems.

Research gaps: Based on the reviewed studies, the existing approaches demonstrate the potential of AI and ML techniques for pharmaceutical manufacturing optimization, quality prediction, fault detection, and process monitoring. There are, however, some drawbacks such as: limited availability of datasets, not enough validation in real-world scenarios, computational complexity, and decreased model transferability from one manufacturing environment to another. Existing literature is mostly devoted to the prediction of dissolution, detection of out of specification samples, prediction of demand, identification of defects, and so on, without the establishment of a complete optimization framework. To fill these research gaps, this work introduces an intelligent pharmaceutical manufacturing optimization model based on the XGBoost algorithm, which can enhance the predictive capability and reliability of the production process by dealing with high-dimensional manufacturing data, class imbalance, and complex production patterns.

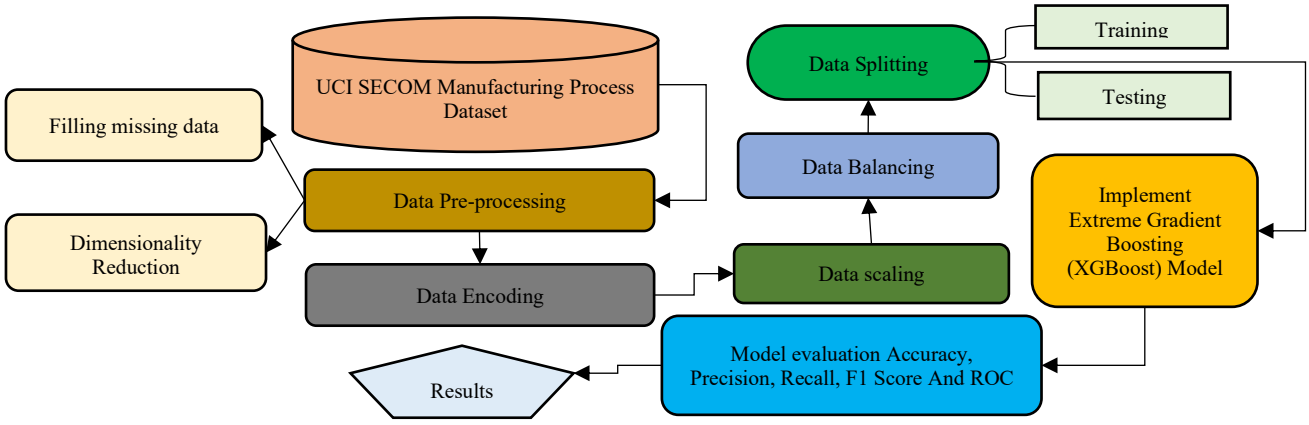

Fig. 1. Proposed flowchart for Intelligent Pharmaceutical Manufacturing Optimization using machine learning

III. RESEARCH METHODOLOGY

The proposed methodology utilizes the UCI SECOM Manufacturing Process Dataset, followed by preprocessing techniques including missing value handling, feature reduction, encoding, scaling, and ADASYN-based data balancing. An intelligent pharmaceutical production optimization model based on XGBoost is suggested, with the processed data divided into a training set and a testing set. ACC, PRE, REC, F1, and ROC curve analysis are employed to assess the model's efficacy. The proposed workflow for intelligent pharmaceutical manufacturing optimization using ML techniques is shown in Fig. 1.

The next section lays out the processes and actions for putting the suggested technique into action:

J. Glob. Res. Multidiscip. Studies, 2 (7):29-36
DOI: https://doi.org/10.5281/zenodo.21836894

A. Dataset Collection and Analysis

The proposed study utilize the UCI SECOM Manufacturing Process Dataset comprising 1567 manufacturing samples, 591 sensor-based process measurements, including 104 failure samples and 1,463 normal samples, suitable for pharmaceutical manufacturing optimization. The distribution, feature correlations etc. are examined with the help of data visualizations like bar plots and heatmaps, which are presented below:

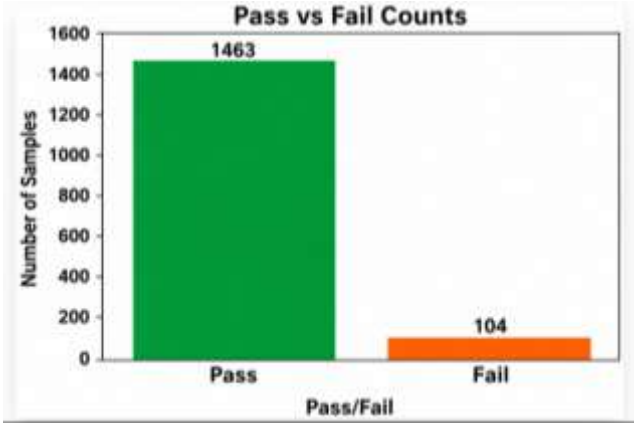

Fig. 2. Bar graph of class distribution

The UCI SECOM Manufacturing Process Dataset distribution of samples is shown in Fig. 2. The data set is highly imbalanced, with 104 fail samples and 1463 pass samples, making it necessary to use suitable preprocessing methods to enhance the performance of the predictive model.

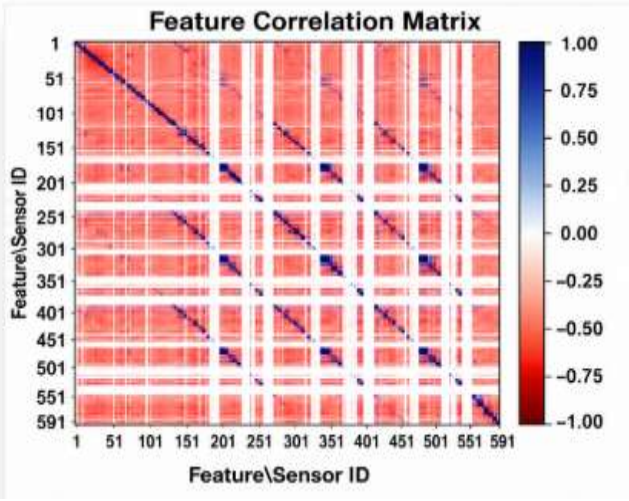

Fig. 3. Feature Correlation Matrix

Fig 3 illustrates the correlation matrix of the sensor features in the UCI SECOM Manufacturing Process Dataset. The heatmap represents the intensity and direction of mutual dependencies between features, which allows the selection of features with a strong correlation to each other, thereby minimizing redundancy when preprocessing data and selecting features.

The pairwise relationships and distributions of selected sensor features (47, 58, 77, and 87) are shown in Fig 4 for the Pass and Fail classes. The scatter and density plot show the separation and correlation between the features, offering the insight of their discriminative power for manufacturing quality prediction.

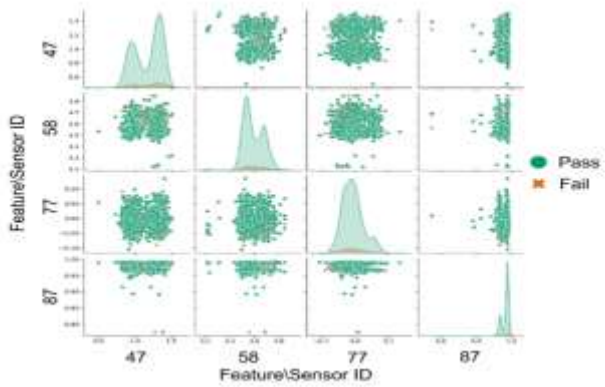

Fig. 4. Feature Distribution of Pass and Fail Classes

B. Data Pre-processing

The UCI SECOM Manufacturing Process Dataset has been carefully preprocessed via a complete pipeline to achieve better data quality and model performance. Data cleaning, imputation of missing values, feature selection, data balancing using ADASYN, data labelling, and feature scaling were the preprocessing stages summarized below:

- **Filling missing data:** Missing data was addressed using an appropriate method to ensure that the data was complete, the quality of the data set was enhanced, and the predictive model was more reliable.
- **Dimensionality Reduction:** This was done to remove redundant and less informative features, which reduced computational complexity, improved model efficiency and enhanced classification performance.

C. Data Encoding

The categorical features were converted into numerical vectors using the One-Hot Encoding approach. Each category was represented by a separate binary feature. This encoding technique avoids ordinal relationships among categories and ensures that ML models can handle categorical attributes effectively and enhance prediction ACC.

D. Data scaling

The data set was made standard by using the Standard Scaler method, which involved adjusting the values of each feature such that the mean (μ) was set to 0 and the standard deviation (σ) was adjusted to 1. This scaling process minimizes the impact of different feature ranges, boosts numerical stability, and optimizes the effectiveness and stability of ML models. Equation (1) shows the mathematical expression for the Standard Scaler.

$$x_{scaled} = \frac{x-\mu}{\sigma} \quad (1)$$

Where σ stands for the feature values' standard deviation and μ for their mean.

E. Data Balancing

Data balancing involves changing the distribution of classes to enhance the performance of the ML model. This is done when there is an imbalance in the dataset, meaning that one class contains much less samples than the others. The ADASYN method was used to create synthetic samples of the minority class in order to correct the dataset's class imbalance. The ML model's generalizability, performance, and class distribution were all improved by this balancing approach, which also decreased prediction bias.

J. Glob. Res. Multidiscip. Studies, 2 (7):29-36
DOI: https://doi.org/10.5281/zenodo.21836894

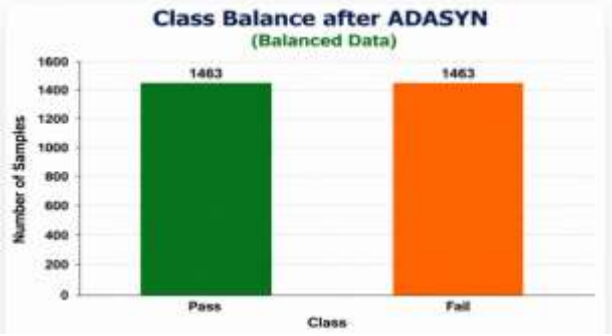

Fig. 5. Balanced Class Distribution after ADASYN

The ADASYN oversampling technique resulted in the balanced distribution of classes seen in Fig 5. The minority Fail class was synthetically expanded to align with the Pass class and a balanced set of 1,463 samples for each class was prepared for successful model training.

F. Data Splitting

There is a 7:3 random split in the first SECOM dataset between training and testing data. This indicates that the training data is represented by 70% of the samples and the testing data by 30%.

G. Proposed Extreme Gradient Boosting (XGBoost) Model

The proposed study is based on the development of a supervised ML framework using the XGBoost algorithm for predictive optimization in intelligent pharmaceutical manufacturing. One ensemble learning approach that uses many DTs models to get good prediction performance is the XGBoost. It aims for optimal outcomes by minimizing a loss function that accounts for the discordance between target and anticipated values. The XGBoost regression model is mathematically represented as follows: (2):

$$y = f(x) \tag{2}$$

XGBoost is a model that takes a vector of input parameters (such square footage, number of bedrooms, etc.) and utilizes them to predict a variable y, where y is the expected property price. Mean squared error (MSE) is a loss function that XGBoost builds a network of decision trees to minimize in order to compute f(x). In order to make a final forecast, the model takes the output of several decision trees and merges them. At its most basic, the XGBoost regression model is (3):

$$y = \sum(k = 1\ to\ K)\ fk(x) \tag{3}$$

The kth decision tree's prediction is shown by $fk(x)$, where K is the number of decision trees in the ensemble. All trees have their anticipated outcomes derived from the weighted average of the values learned for their leaves during training. Calculate the XGBoost model's prediction for a given input x by adding the predictions of all the ensemble decision trees. The suggested XGBoost model was built using the following optimization hyperparameters: The following parameters are set: estimators = 200, learning rate = 0.05, max_depth = 6, min_child_weight = 1, subsample = 0.8, colsample_bytree = 0.8, gamma = 0, and random state = 42. The model used the logistic loss function with the tree-based booster (gbtree) for binary classification of manufacturing outcomes. These parameter settings improved model stability, reduced overfitting, and enhanced prediction ACC for intelligent pharmaceutical manufacturing optimization.

H. Evaluation Metrics

The confusion matrix was used to calculate the F1, ACC, PRE, and REC, which were then used to evaluate the model's performance. The confusion matrix summarizes the outcomes of predictions in terms of TP, TN, FP, and FN, which is important for evaluating the accuracy of a classification system's ACC. These measures summarize the degree of ACC in the classification of positive and negative cases without misclassification. Together they provide an effective and robust account of how the proposed classification model works, is stable, and is applicable. Equations (4)–(7) display key performance indicators based on these values, such as ACC, PRE, REC, and F1.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{2}$$

$$\text{PR} = \frac{\text{TP}}{\text{TP+FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

The suggested classification model was evaluated using ACC, PRE, REC, F1, and ROC curve. The ACC measures the accuracy of all instances categorized, PRE measures the accuracy of positive predictions, and REC measures the number of right positive predictions. An F1-balanced score is the harmonic mean of two scores, PRE and REC. Additionally, the ROC curve provides a comprehensive view of the discriminative capacity of the model at different classification threshold values and illustrates the trade-off between TPR and FPR.

IV. Results and Discussion

This section details the experimental setup and performance analysis, which demonstrate the computational efficiency, accuracy, and efficiency of the proposed model. An HP Z2 Mini G9 Workstation, which comes with a 12th Gen Intel Core i9-12900 CPU, 64 GB of RAM, and an NVIDIA RTX A2000 12 GB GPU, was used for the tests and model prediction. Python 3.10.12, NumPy 2.2.6, scikit-learn 1.7.2, pandas 2.3.3, and other apps were utilized. For the sake of reproducibility, and did all of the tests on the identical set of hardware and software.

TABLE II. Performance Evaluation Results of the Proposed Model for Intelligent Pharmaceutical Manufacturing Optimization

Matrix	proposed Extreme Gradient Boosting (XGBoost) Model
Accuracy	98.93
Precision	98.97
Recall	98.99
F1-score	98.92

The UCI SECOM Manufacturing Process Dataset was used to train and assess the proposed XGBoost model with performance metrics like ACC, PRE, REC, and F1. The suggested model achieved an ACC of 98.93%, PRE of 98.97%, REC of 98.99%, and F1 of 98.92%, which indicated that the model has good predictive power and efficacy in intelligent pharmaceutical production optimization, as seen in Table II.

J. Glob. Res. Multidiscip. Studies, 2 (7):29-36
DOI: https://doi.org/10.5281/zenodo.21836894

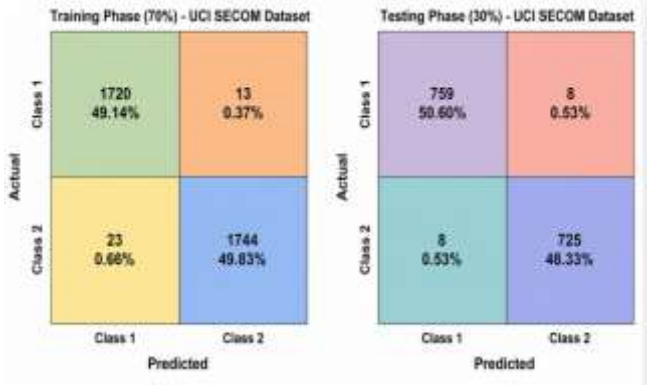

Fig. 6. Confusion Matrix for Training and Testing Phases for the proposed model

The proposed model's classification performance on the UCI SECOM dataset with 70% training data and 30% testing data, as illustrated in Fig 6, the Confusion Matrix for the Training and Testing Phases. Numbers that are tiny away from the diagonal indicate a low misclassification rate, while numbers that are significant near the diagonal indicate that most samples were assigned properly. The model's superb classification ability, great learning capacity, and ability to generalize to unknown data are all supported by the excellent results that were produced.

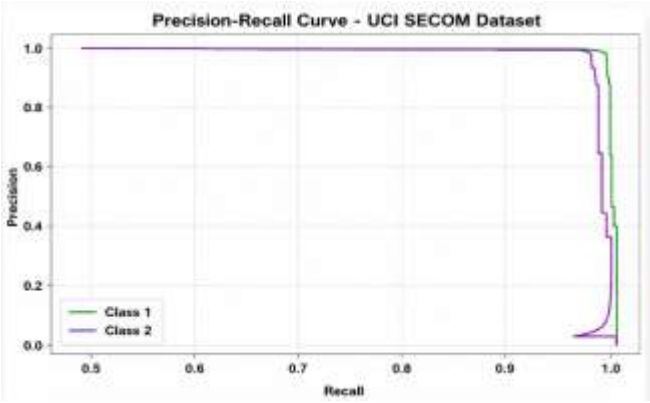

Fig. 7. Precision–Recall Curve for the Proposed Model

Fig 7 depicts PRE-REC Curves for both classes of the SECOM UCI dataset. The curves are aligned near the upper right corner for all the REC values, suggesting that the proposed model has high PRE and excellent REC. This performance shows good classification ability, good discrimination between classes, and good predictive performance with a low false positive and low false negative rate.

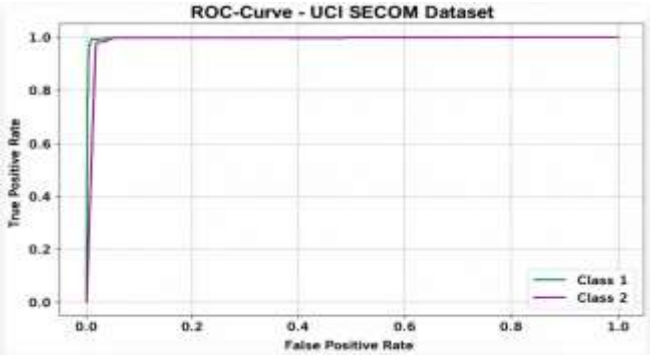

Fig. 8. Receiver Operating Characteristic (ROC) Curve for the Proposed Model

Fig 8's ROC Curve displays the suggested classification model's performance for both classes with respect to the TPR and the FPR. Their strong AUC scores and proximity to the top left corner of the ROC curves indicate that they can distinguish well between the two groups. The findings show that the proposed model performs well on the UCI SECOM data set for classification, has low false positive rates, and achieves extremely high sensitivity.

A. Comparative Analysis

Table III compares the proposed model's ACC to that of other models already in existence in order to evaluate the model's performance. This section compares the suggested XGBoost model's performance against that of other ML and DL methods that use ACC, PRE, REC, and F1 evaluation criteria. At 98.93%, MLP 97.6%, DNN 92%, DT 88.1%, RF 83%, and SVM 71%, the suggested XGBoost model is the most effective. The effectiveness of the proposed model in learning complex patterns from pharmaceutical manufacturing data, and enhancing the ACC of the prediction, is illustrated in the superior performance. The obtained results confirm the potential of the proposed XGBoost method in intelligent optimization of pharmaceutical manufacturing.

TABLE III. PERFORMANCE COMPARISON OF VARIOUS MACHINE LEARNING AND DEEP LEARNING MODELS FOR PHARMACEUTICAL MANUFACTURING OPTIMIZATION

Model	Accuracy	Precision	Recall	F1-score
SVM[29]	71	50	59	-
RF[30]	83	50	26	22
DT[31]	88.1	30.1	09.6	93
DNN[32]	92	90.3	92	91.1
MLP[33]	97.6	95.4	99.9	97.6
Proposed XGBoost	98.93	98.97	98.99	98.92

The proposed XGBoost model had ACC as 98.93%, which is superior in predicting intelligent pharmaceutical manufacturing optimization. The model is able to extract complex patterns and relationships from manufacturing data, allowing for accurate prediction and optimization of manufacturing processes. High ACC and a learning ability are recognized as effective in enhancing the efficiency of manufacturing, product quality, and data-driven decision-making processes in pharmaceutical industries.

V. CONCLUSION AND FUTURE STUDY

The use of AI and ML is revolutionizing the healthcare business, especially in the areas of medicine development and distribution. Particularly in the areas of medication administration and discovery, AI and ML are causing a paradigm shift in the pharmaceutical sector. By analyzing massive volumes of data, these technologies can personalize formulations and predict how patients will respond to PRE medication. Drug delivery systems made with nanoparticles are improved by AI models, which boost stability, bioavailability, and targeting. Additionally, ML enables the monitoring and adapting the release of the drug during the experiment, leading to improved therapeutic results. The proposed XGBoost model gave better ACC of 98.93%, which is better than that of the existing models like SVM (71%), RF (83%), DT (88.1%), DNN (92%), and MLP (97.6%) as per the experimental results. The results obtained demonstrate the effectiveness and reliability of the proposed approach for optimizing the pharmaceutical manufacturing process, by

increasing the ACC of the prediction, process monitoring and efficiency of decision-making.

The framework proposed by XGBoost is only applicable to one benchmark dataset, which may not cover the actual spectrum of pharmaceutical manufacturing environments. The study also does not take into account real-time deployment, model interpretability and integration with an automated production system. Future work will be extended by using larger multi-site datasets and by integrating Explaining AI and develop real-time adaptive manufacturing optimization solutions.

- **Funding:** This research received no external funding.
- **Institutional Review Board Statement:** Not applicable.
- **Informed Consent Statement:** Not applicable.
- **Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author upon reasonable request.
- **Conflicts of Interest:** The author declares no conflict of interest.

How to cite this article:
C, Kumar Patel, "Artificial Intelligence-Assisted Pharmaceutical Manufacturing Optimization Using Predictive Machine Learning Techniques," *Journal of Global Research in Multidisciplinary Studies*, vol. 2, no. 7, pp. 29–36, Jul. 2026., Doi 10.5281/zenodo.21836894

REFERENCES

[1] S. Mahmud, K. K. Mohammed, V. R. Jain, and S. A. Shah, "Artificial Intelligence-Driven Predictive Models for Identifying Risk Factors of Chronic Diseases," in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, Boston, MA, USA: IEEE, Mar. 2026, pp. 01–06. doi: 10.1109/ISDFS69419.2026.11459003.

[2] A. Warrier and A. K S, "Autonomous Agentic AI for Clinical Workflow Orchestration: Self-Managing Healthcare Operations," in *2025 6th International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, IEEE, Dec. 2025, pp. 1450–1457. doi: 10.1109/ICICNIS66685.2025.11315712.

[3] R. Kant, R. Rao Thallada, B. Pandey, and P. Srivastava, "AI-Based Cybersecurity in Healthcare: A Data-Driven, Governance-Aware Framework for Secure Clinical Systems," in *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, 2026, pp. 1–5. doi: 10.1109/ICAIC67076.2026.11395836.

[4] N. N. Agarwal *et al.*, "Technological Forecasting & Social Change E-health and wellbeing monitoring using smart healthcare devices : An empirical investigation," *Res. Policy*, 2020.

[5] 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.

[6] S. Kukar, "Explainable AI Models for Transparent and Ethical Customer Relationship Management Decision," in *2026 IEEE 2nd International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, Houston, TX, USA: IEEE, Mar. 2026, pp. 1–7. doi: 10.1109/SATC69565.2026.11542554.

[7] S. Mukherjee, "AI-Driven Deep Learning System for Alzheimer's Disease (AD) Detection Via MRI Imaging," in *2026 International Conference on Intelligent Computing, Networks, and Security (IC-ICNS)*, Bhubaneswar, India: IEEE, 2026, pp. 1–6, March. doi: 10.1109/IC-ICNS68863.2026.11537765.

[8] V. K. Sharma, Z. Khan, R. K. Sharma, and D. Jindal, "Artificial Formulation Development in Pharmaceutics : Current Applications , Challenges and Future Perspectives," vol. 4, no. 7, pp. 193–198, 2026, doi: 10.5281/zenodo.21113149.

[9] S. R. Katta, "Optimizing Pharmaceutical Manufacturing through AI-Powered Calibration and Maintenance Workflow Management," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 1, no. 4, pp. 2175–2181, Nov. 2023, doi: 10.51219/JAIMLD/Srikanth-reddy-katta/476.

[10] O. Afzal *et al.*, "Nanoparticles in Drug Delivery: From History to Therapeutic Applications," *Nanomaterials*, vol. 12, no. 24, p. 4494, Dec. 2022, doi: 10.3390/nano12244494.

[11] Y. Jia *et al.*, "Approved Nanomedicine against Diseases," *Pharmaceutics*, vol. 15, no. 3, p. 774, Feb. 2023, doi: 10.3390/pharmaceutics15030774.

[12] R. Lingam, "Agentic AI and Federated Learning for Privacy-Preserving Dental Imaging," in *2026 Innovations in Machine, Engineering, and Digital Conference (IMED)*, Kota Kinabalu, Malaysia: IEEE, 2026, pp. 1–6, March. doi: 10.1109/IMED68921.2026.11484887.

[13] A. Warrier, "Real-Time AI Integration Architectures for HIPAA-Compliant Healthcare Data Interoperability," in *International Journal of Emerging Trends in Computer Science and Information Technology*, Eureka Vision Publication, 2025, pp. 74–81. doi: 10.63282/3050-9246/WCAI25-128.

[14] Noorain, V. Srivastava, B. Parveen, and R. Parveen, "Artificial Intelligence in Drug Formulation and Development: Applications and Future Prospects," *Curr. Drug Metab.*, vol. 24, no. 9, pp. 622–634, Sep. 2023, doi: 10.2174/0113892002265786230921062205.

[15] S. Kolluri, J. Lin, R. Liu, Y. Zhang, and W. Zhang, "Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: a Review," *AAPS J.*, vol. 24, no. 1, p. 19, Jan. 2022, doi: 10.1208/s12248-021-00644-3.

[16] P. Suriyaamporn, B. Pamornpathomkul, P. Patrojanasophon, T. Ngawhirunpat, T. Rojanarata, and P. Opanasopit, "The Artificial Intelligence-Powered New Era in Pharmaceutical Research and Development: A Review," *AAPS PharmSciTech*, vol. 25, no. 6, p. 188, Aug. 2024, doi: 10.1208/s12249-024-02901-y.

[17] W. Wang, Z. Ye, H. Gao, and D. Ouyang, "Computational pharmaceutics - A new paradigm of drug delivery," *J. Control. Release*, vol. 338, pp. 119–136, Oct. 2021, doi: 10.1016/j.jconrel.2021.08.030.

[18] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, and C. Allen, "Machine learning directed drug formulation development," *Adv. Drug Deliv. Rev.*, vol. 175, p. 113806, Aug. 2021, doi: 10.1016/j.addr.2021.05.016.

[19] S. A. Devanira Poovaiah, M. Sankaran, G. Namazov, N. Jooluri, and M. Mukhlisa, "Alzheimer's Disease Detection from MRI Scans using Deep Hybrid Architecture with Metaheuristic Hyperparameter Tuning," in *2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)*, IEEE, Jul. 2025, pp. 1–7. doi: 10.1109/ISAC364032.2025.11156608.

[20] J. Kehrein, A. Bunker, and R. Luxenhofer, "POxload: Machine Learning Estimates Drug Loadings of Polymeric Micelles," *Mol. Pharm.*, vol. 21, no. 7, pp. 3356–3374, Jul. 2024, doi: 10.1021/acs.molpharmaceut.4c00086.

[21] N. Mozafari, N. Mozafari, A. Dehshahri, and A. Azadi, "Knowledge Gaps in Generating Cell-Based Drug Delivery Systems and a Possible Meeting with Artificial Intelligence," *Mol. Pharm.*, vol. 20, no. 8, pp. 3757–3778, Aug. 2023, doi: 10.1021/acs.molpharmaceut.3c00162.

[22] K. Thupphae and N. Wiwatwattana, "Data-Driven Application and Comparison of Machine Learning Models for Drug Dissolution Prediction in Pharmaceutical Quality Risk Management," in *2026 23rd International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, Jun. 2026, pp. 144–149. doi: 10.1109/JCSSE68839.2026.11596766.

[23] R. Mannam, "A Machine Learning Framework to Predict QC Lab OOS Events and Reduce Rework," in *2026 5th International Conference on Computer Networks, Big Data and IoT (ICCBI)*, IEEE, Jun. 2026, pp. 558–564. doi: 10.1109/ICCBI68589.2026.11619806.

[24] R. M, R. Patil, D. G. Sarode, and S. A. Sakharwade, "AI-based Smart Control Systems for Induction Brazing in Industrial Applications," in *2025 5th International Conference on Trends in Material Science and Inventive Materials (ICTMIM)*, IEEE, Apr. 2025, pp. 1530–1537. doi: 10.1109/ICTMIM65579.2025.10988032.

[25] M. Bhende, A. Salati, P. S. Kumar, P. Senthil, S. M, and G. Ali, "The Role of AI and Robotics in Fully Autonomous Pharmaceutical Manufacturing Facilities," in *2025 IEEE 4th International Conference for Advancement in Technology (ICONAT)*, IEEE, Sep. 2025, pp. 1–7. doi: 10.1109/ICONAT66879.2025.11362620.

[26] F. Kakavandi, P. Han, T. B. Nerenst, R. de Reus, C. Gomes, and P. G. Larsen, "Comparative Study in Fault Detection under Data Scarcity for Pharmaceutical Device Manufacturing," in *2025 International Conference on Control, Automation and Diagnosis (ICCAD)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/ICCAD64771.2025.11099417.

[27] P. Malathi, S. Jansi, M. V. Prabhakaran, K. Senbagam, M. P. Sujatha, and R. Sadaieswaran, "Big Data Analytics for Quality Control in The Pharmaceutical Industry," in *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, Mar. 2024, pp. 669–674. doi: 10.1109/ICACCS60874.2024.10717314.

[28] L. D. Lam, B. P. Le Luong, H. T. Mai Linh, and P. M. Hung, "Application of Machine Learning in Predicting the Amount of Pharmaceutical Drugs Ordered for the Manufacturer," in *2023 1st International Conference on Health Science and Technology (ICHST)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICHST59286.2023.10565367.

[29] A. Farrag, M.-K. Ghali, and Y. Jin, "Rare Class Prediction Model for Smart Industry in Semiconductor Manufacturing," 2024.

[30] D. K. Nguyen, C. L. Chan, and D. Van Phan, "Gradient Deep Learning Boosting and Its Application on the Imbalanced Datasets Containing Noises in Manufacturing," *Smart Innov. Syst. Technol.*, vol. 314, pp. 225–235, 2023, doi: 10.1007/978-3-031-05491-4_23.

[31] H. J. Park, Y. S. Koo, H. Y. Yang, Y. S. Han, and C. S. Nam, "Study on Data Preprocessing for Machine Learning Based on Semiconductor Manufacturing Processes," *Sensors*, vol. 24, no. 17, pp. 1–14, 2024, doi: 10.3390/s24175461.

[32] D. Moldovan, I. Anghel, T. Cioara, and I. Salomie, "Particle Swarm Optimization Based Deep Learning Ensemble for Manufacturing Processes," in *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, IEEE, Sep. 2020, pp. 563–570. doi: 10.1109/ICCP51029.2020.9266269.

[33] R. Alharbi, "From synthetic data to actionable insights: fault detection in industry," *Complex Intell. Syst.*, vol. 12, no. 8, 2026, doi: 10.1007/s40747-026-02323-2.