PREDICTION OF API IMPURITIES USING ARTIFICIAL INTELLIGENCE

Ms. Dhanashri Bhonde ,Dr. S.G.Jawarkar, Ms. Eshika Deshmukh , Ms. Diksha Patil

Vidyabharati college of pharmacy, Amravati

ABSTRACT

Controlling impurities in Active Pharmaceutical Ingredients (APIs) remains a key challenge due to complex synthesis routes, diverse contamination sources, and limitations of traditional analytical methods. With increasing demand for safer medicines, Artificial Intelligence (AI) is becoming a valuable tool for predicting, identifying, and monitoring impurities more efficiently. Techniques such as machine learning, deep learning, QSAR modeling, and reaction-prediction algorithms are now integrated with analytical platforms like HPLC, LC–MS, GC–MS, NMR, and PAT to improve impurity profiling. AI can analyze large datasets, detect hidden patterns, and predict degradation pathways and genotoxic risks in advance.

Applications including retention-time prediction, ICH M7-based mutagenicity assessment, and AI-assisted stability studies demonstrate its growing practicality. Despite challenges such as limited data quality, model transparency, and regulatory adaptation, the integration of AI with automation and digital technologies offers strong potential for real-time impurity control and improved quality-by-design approaches in pharmaceutical development.

Keywords: Impurities, API, QSAR, Machine learning, AI, ICH guidelines, Impurity profiling,Impurity prediction

INTRODUCTION

Overview of Active Pharmaceutical Ingredients (APIs)

Active Pharmaceutical Ingredients (APIs) are the parts of a drug that actually do the healing or treating. These can be made in a lab, made from living things, or created through special biological processes. They need to be made carefully and tested a lot to make sure they are right in terms of what they are, how pure they are, how strong they are, and how well they last. Even small mistakes in making an API can affect how well the drug works or if it is safe.

APIs are the actual chemicals in a medicine that make it work.

They can be made by mixing chemicals, taken from nature, grown using bacteria, or made using high-tech biology, like changing DNA. The quality of the API is super important because it affects how safe and effective the final medicine is. That's why it's crucial to follow strict rules when making APIs, like Good Manufacturing Practices (GMP), and to meet rules set by groups like the FDA, EMA, and ICH. This helps make sure each batch of API is the same, works the same, and is safe.

Importance of Impurity Profiling and Control

Impurities are any substances found in the active pharmaceutical ingredient (API) or finished drug product that are not the intended drug substance or additives. Impurity profiling involves carefully finding, identifying, measuring, and controlling these unwanted substances, and it is vital for ensuring patient safety, getting regulatory approval, achieving consistent effectiveness, and maintaining proper documentation of quality and intellectual property. Regulatory guidelines from organizations like ICH, EMA, and FDA set rules for identifying and classifying impurities and require detailed impurity profiles for products that are meant to be sold. This data helps evaluate the risk of toxicity and supports the setting of acceptable limits for impurities. Not properly profiling impurities can result in harmful side effects, product recalls, or rejection by regulators.

Importance of impurity Profiling:

- Even small amounts of impurities can be harmful, especially if they are genotoxic.
- Impurities can affect how stable a drug is, how well it is absorbed by the body, and its effectiveness.
- Getting approval from regulators requires full understanding and justification of impurity levels.

• Contaminants have caused serious problems in the pharmaceutical industry, such as the valsartan NDMA impurity incident.

Analytical methods like HPLC, UHPLC, GC-MS, LC-MS/MS, NMR, and ICP-MS are key tools, but new impurity challenges need more advanced, automated, and efficient testing solutions.

Challenges in Impurity Prediction and Detection

Impurity analysis presents both practical and conceptual difficulties. Impurities often appear in very small amounts, such as parts per million or even parts per billion, and they can be of many different chemical types, including organic breakdown products, synthetic by-products, elemental impurities, leftover solvents, and genotoxic substances. These impurities can form at various stages, such as during the manufacturing process, purification, formulation, storage, or transportation. Complex formulations, especially those involving biologics or combination products, can make things more complicated by increasing matrix effects and making it harder to detect impurities.

Analytical challenges include separating compounds that come out of the system at the same time, achieving sufficient sensitivity and selectivity, identifying unknown substances, and understanding how impurities form. Predicting which impurities might form like through degradation pathways or process-related by-products is difficult because it requires a deep understanding of chemistry, process conditions, excipients, and stress factors. Although traditional methods like HPLC, GC, MS, and NMR are powerful, they can be time consuming and often need expert-level method development. Even with advanced analytical technology, impurity profiling is one of the most difficult areas in pharmaceutical quality control.

Need for Artificial Intelligence (AI) in Pharmaceutical Quality Control

AI and machine learning provide tools to tackle several challenges in impurity science. They can look at big, mixed datasets like analytical results, process settings, stability data, and reaction pathways from literature to spot patterns that humans might miss. These tools can help predict possible impurities or degradation products, decide which substances to focus on, speed up method development by suggesting things like chromatographic settings, and support real-time quality control when used with PAT (process analytical technology) and digital sensors. Some AI uses already in place include models that predict how impurities form, help break down and classify complex spectra, find unusual patterns in quality control data, and improve analytical conditions. While AI can't replace detailed structural analysis, it can cut down on experiments, guide efforts toward high-risk impurities, and boost compliance with regulations if properly tested and documented. Important steps for using AI include having good quality data, making sure the AI's decisions are clear and understandable for regulatory approval, ensuring it works well through testing, and connecting it smoothly with current quality systems. Artificial Intelligence (AI) brings significant opportunities for analysing Impurities, predicting their actions, and improving quality control in analysis.

AI is especially helpful in:

- Predicting impurities based on molecular structures
- Looking for trends in impurities during stability testing
- Finding issues in quality control data
- Simulating how reactions go during drug manufacturing
- Creating better strategies for purifying compounds

Regulatory agencies are starting to recognize AI's value, as long as it is clear, explainable, and properly validated.

CLASSIFICATION

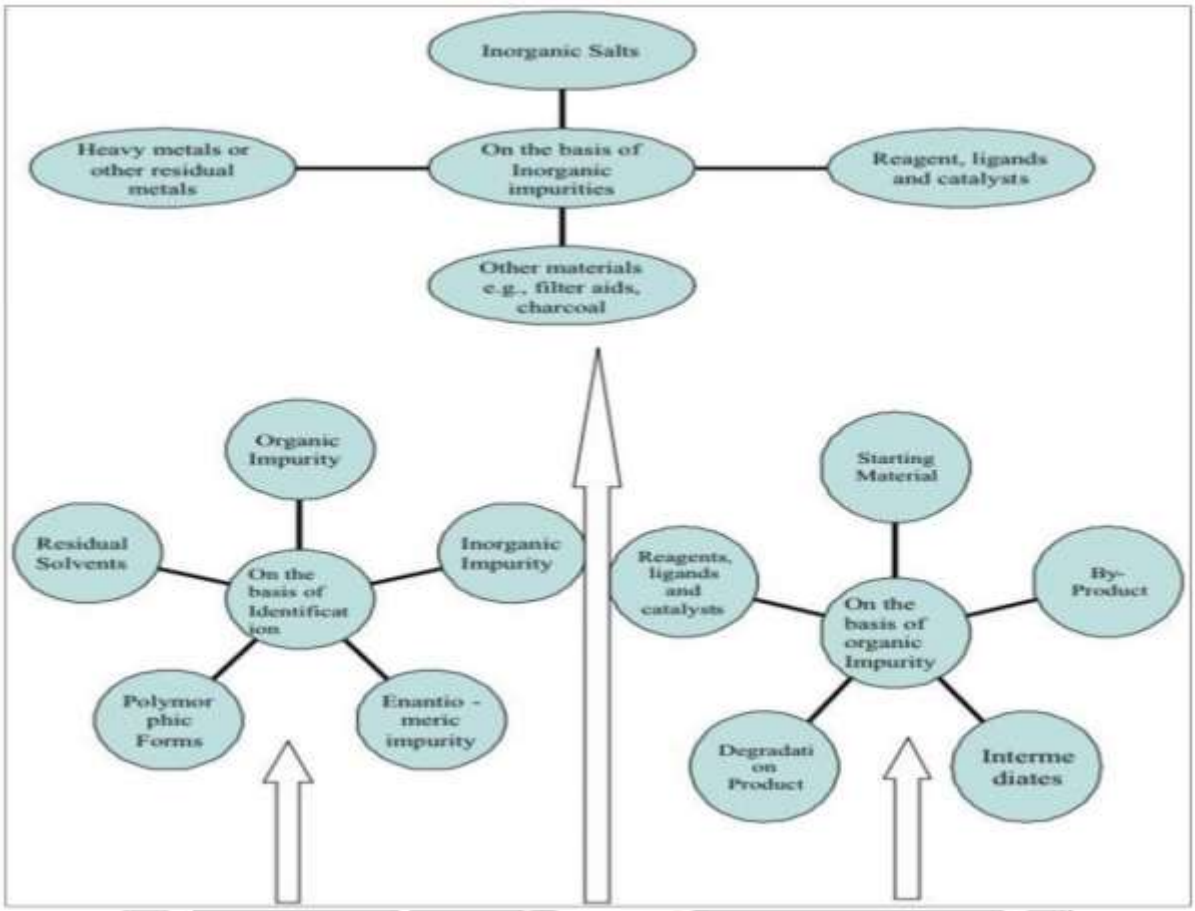

Fig no.1: Source of impurity

Source of Impurities

Impurities are categorized in different ways by different authorities.

According to the United States Pharmacopoeia:

a. Impurities found in official medicines.

b. Common impurities.

c. Organic volatile impurities.

According to ICH:

a. Organic impurity (from the manufacturing process or the drug itself).

b. Inorganic impurity.

c. Residual solvents.

REGULATORY GUIDELINES

ICH Q3A(R2): Impurities in New Drug Substances (APIs)

ICH Q3A outlines the impurity thresholds and reporting standards for organic, inorganic, and residual impurities in APIs. AI-driven predictive models must adhere to these guidelines to maintain compliance.

ICH Q3B(R2): Impurities in New Drug Products

Q3B addresses impurities in the final formulation, not the API. AI systems predicting degradation pathways in drug products must adhere to Q3B limits.

ICH Q3C(R8): Residual Solvents

Q3C classifies residual solvents and establishes permissible daily exposure (PDE) values.AI models predicting solvent retention or related impurity formation must operate within Q3C safety limits.

AI IN PHARMACEUTICAL IMPURITY ANALYSIS

Basics of AI &ML

Artificial Intelligence (AI) and Machine Learning (ML) assist in analysing extensive and complex impurity data in the pharmaceutical industry.

ML algorithms learn from historical experimental data, chromatographic results, and chemical structures to predict potential impurities, classify impurity types, and aid decision-making. These techniques reduce manual efforts and enhance the accuracy of impurity detection.

Common AI Models Used

- Supervised Learning Models (e.g., Random Forest, Support Vector Machines) – used for impurity classification and quantification.
- Unsupervised Models (e.g., K-means clustering, PCA) – useful for pattern recognition, grouping unknown impurities, and detecting outliers.
- Deep Learning Models (e.g., neural networks, CNNs) – applied for spectral analysis, chromatogram interpretation, and predicting impurity formation pathways.
- QSAR Models – predict impurity formation based on chemical structure features.

PREDICTIVE MODELING OF IMPURITIES

What is predictive modelling

Predictive modelling is a method that uses statistical, computational, or artificial intelligence (AI) to forecast how impurities form, their levels, and how they behave in Active Pharmaceutical Ingredients (APIs).

It uses past data, reaction details, how substances break down, and test results to build models that can predict impurity levels before or during the making of the drug. In pharmaceutical quality control, predictive modelling helps find possible impurities early, reduces the need for many experiments, and makes the production process more reliable.

a. Data Collection

Gather data from experiments, including reaction conditions, starting materials, solvents, catalysts, pH, temperature, time, and test results from methods like HPLC, LC–MS, and GC–MS.

- Clean the data to get rid of mistakes, errors, and odd results.
- The quality of the predictive model depends a lot on how accurate and complete the data is.

b. Model Development

AI methods such as:

- Machine learning models: Random Forests, Support Vector Machines, Neural Networks
- Chemoinformatic tools: QSAR, QSPR, reaction modelling.
- Mechanistic models: kinetic models, degradation pathway modelling.
- The model learns from the data to understand how impurities form under various conditions.

c. Model Validation

Check the model using test data, cross-validation, or external testing.

Use performance measures like:

R^2 (coefficient of determination)

• RMSE (root mean square error)

- Accuracy, sensitivity, specificity (for classification)
- A validated model ensures reliable impurity predictions during large-scale production and manufacturing.

AI with Analytical Techniques

AI improves impurity predictions when used with modern analytical tools:

AI + HPLC/LC-MS: Helps find unknown impurities, predict how long they stay in the system, and match chromatographic patterns to impurity formation.

AI + Spectroscopy (NMR, IR, UV): Helps spot early signs of breakdown and predicts how similar degradation products might look.

AI + Process Analytical Technology (PAT): Enables real-time monitoring of impurities using data from Raman, NIR, or MS sensors during production.

AI + Stability Studies: Predicts how stable an API is and how impurities grow under ICH conditions (like accelerated aging).

Overall, combining AI with analytical methods improves speed, sensitivity, early detection, and how accurate predictions are, supporting better control of impurities in APIs.

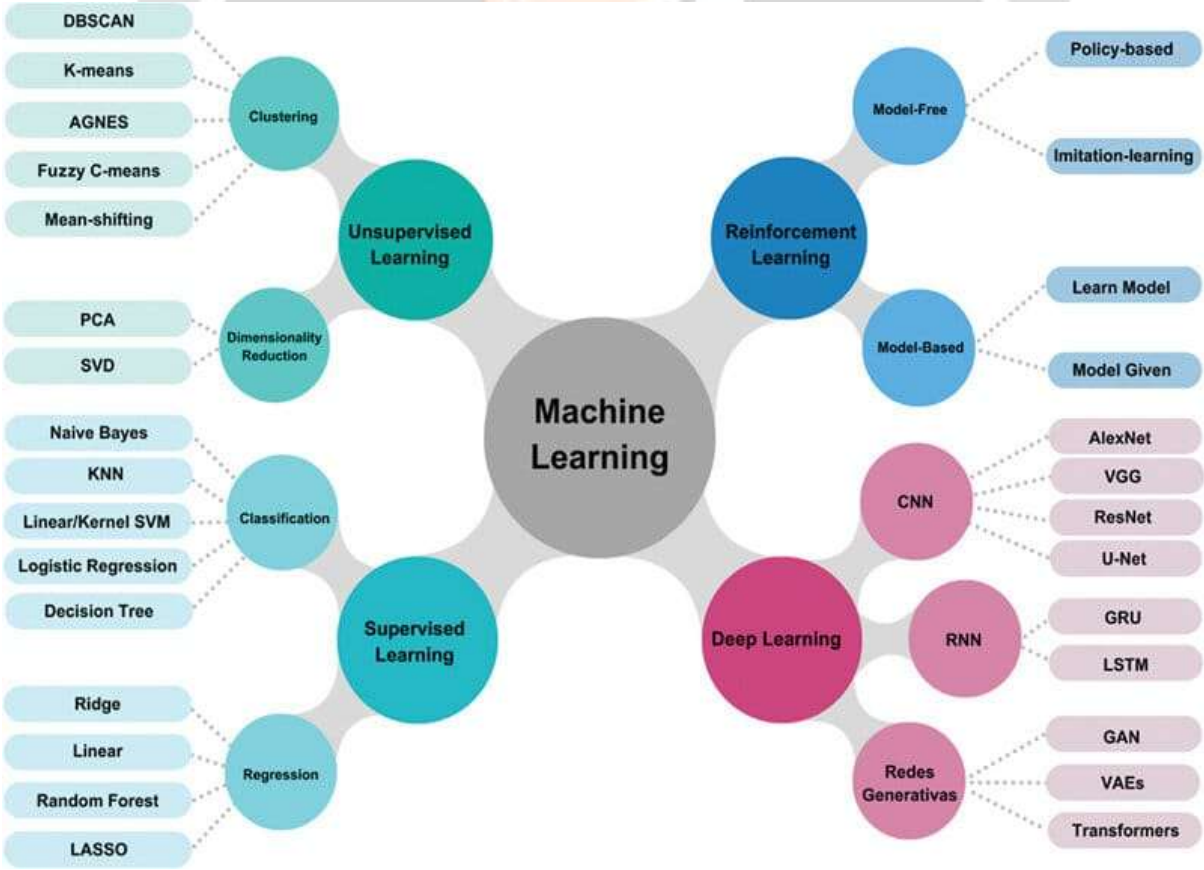

Fig no.2: Machine Learning

AI METHODS IN PREDICTIVE MODELING OF API IMPURITIES

Machine Learning Methods

Machine learning (ML) is commonly used to predict impurities in active pharmaceutical ingredients (APIs) because it can find complex relationships between process factors, reaction conditions, and how impurities form. Common ML models include Random Forests, Support Vector Machines (SVM), k-Nearest Neighbours (kNN), Gradient Boosting, and Ensemble models.

These models help with:

- Predicting impurity levels using past manufacturing data
- Finding important quality attributes and process parameters
- Cutting down on experiments by testing different process conditions virtuall
-
- ML works well for predicting impurities because it handles complicated chemical data and fits with analytical data like HPLC, MS, NMR, and PAT.

Deep Learning

Deep learning (DL) models, such as artificial neural networks (ANNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs), are used when there is a lot of data available.

DL is helpful for:

- Predicting how APIs break down
- Interpreting spectral data (like NMR, MS, and IR)
- Modelling impurity formation that follows complicated patterns
- Finding hidden patterns in big chemical datasets

DL usually gives better results than traditional ML, but it needs more data and stronger computer power. In impurity prediction, DL is often used together with chemometric or spectroscopy methods.

QSAR (Quantitative Structure–Activity Relationship) Models

QSAR models use the chemical structure of APIs, intermediates, or breakdown products to predict impurity formation or how reactive they are. They rely on molecular features like electronic, steric, and topological properties to model:

- Toxicity of impurities
- Likelihood of impurity creation
- How impurities break down
- Alerts for dangerous impurities (like nitrosamines)

QSAR is especially important for checking dangerous impurities and is recommended in rules for predicting toxicity before running tests in the lab.

Software and Platforms Used

AI-driven impurity prediction depends on computational tools that help with data preparation, building models, and creating visualizations

MATLAB

This tool is used for chemometrics and doing complex data analysis with multiple variables.

- Used with chemometric methods like PCA, PLS, and MLR.
- Good for analyzing impurities using techniques like NIR, UV, and Raman.

Python

python is the main programming environment because of its open-source libraries:

- scikit-learn → Provides machine learning tools for predicting impurities.
- TensorFlow/Keras, PyTorch → Used for deep learning with spectral and chromatographic data.
- RDKit → Helps calculate molecular features.
- XGBoost → Offers high accuracy for predicting impurities in pharmaceutical data.

Python allows building complete pipelines for predicting impurities.

KNIME

This is a platform where you can drag and drop tools to work on data.

- Offers drag-and-drop workflows that don't require coding
- Can include cheminformatics plugins.
- Used for creating QSAR models, predicting toxicity, and classifying impurities

TensorFlow and PyTorch

Deep learning frameworks are used for:

- Predicting the results of forced degradation tests.
- Identifying unknown peaks in spectra.
- Automatically predicting the structure of impurities from mass spectrometry data.

CASE STUDIES: EXAMPLES WHERE AI PREDICTED IMPURITIES

AI-Assisted Prediction of Process-Related Impurities in Small-Molecule Synthesis

Machine learning models have been developed to map reaction conditions to impurity formation outcomes .In one notable study, supervised learning algorithms such as random forests and gradient boosting were trained on historical process development data to identify impurity formation pathways in small-molecule APIs. These models successfully predicted which by-products would appear under various parameters like temperature, solvent, pH, and catalyst loading. This enabled chemists to optimize reaction conditions early, reducing the need for costly laboratory experiments.

Deep Learning Models for Predicting Degradation Impurities

Deep neural networks have been applied to accelerate stability studies by predicting likely degradation products under stress conditions (e.g., humidity, heat, light). Using structural information (SMILES) and reaction rules, deep learning frameworks predicted probable degradation pathways for antibiotics and anticancer compounds. The models helped identify photolytic and hydrolytic impurities in accelerated stability testing, enabling targeted monitoring during the development phase.

QSAR Models for Mutagenic Impurity Prediction (ICH M7)

QSAR-based AI tools—like Derek Nexus and Lead scope —have been used by pharmaceutical companies to predict mutagenic impurities in drug substances. These systems use large toxicology databases and structural alerts to classify potential impurities as Class 1–5 under ICH M7. In a real-world case, a company used QSAR predictions to identify a potential mutagenic impurity early in process design. Reformulating the synthetic route helped eliminate this impurity before production scale-up

Retention-Time Prediction to Identify Unknown Impurities (LC–MS)

AI models have been used to predict chromatographic retention times of impurities in LC–MS workflows. In a case involving complex API matrices, a neural-network retention-time predictor matched unknown chromatographic peaks with likely impurity structures generated by in-silico libraries. This drastically reduced the time required for structure elucidation and enabled rapid identification of trace-level impurities.

Integration of AI With PAT (Process Analytical Technology) to Predict Real-Time Impurities

A big pharmaceutical company used AI models that were trained using data from PAT (Process Analytical Technology), like NIR and Raman, to predict impurity levels as the chemical reaction happened. The AI system could let them know when impurity levels went beyond safe limits, so they could quickly fix the process.

Sensitivity for low-level impurities — Traditional: depends on the equipment and method used; AI: can detect impurities better through smart algorithms, but the detection limits still need to be approved.

Structural elucidation — Traditional: needs manual work with MS/MS and NMR; AI: helps suggest possible structures faster, but still needs proof from experiments.

Regulatory readiness — Traditional: has well-established and accepted methods; AI: is a useful tool but needs to be explained clearly and proven reliable.

ADVANTAGES OF AI IN IMPURITY PREDICTION

Higher Accuracy

AI models, especially machine learning (ML) and deep learning (DL), can look at big sets of chemical, analytical, and reaction data to find complex and hard-to-spot patterns that traditional methods might miss.

They help improve accuracy by:

- Learning from past impurity data
- Finding small but important connections between reaction conditions and how impurities form
- Cutting down on mistakes that people can make when figuring out structures or reading chromatography results

Studies show that using AI helps predict impurity structures, mutagenicity, and how impurities move in chromatography more precisely, which makes quality control better.

Faster Prediction

AI speeds up impurity detection and forecasting by:

- Automatically analyzing LC–MS/MS data
- Guessing possible impurities from chemical pathways in just minutes
- Quickly checking thousands of molecules for potential risks
  This saves time on experiments and helps make decisions faster during drug development.

Cost Reduction

AI lowers costs in analysis and development by:

- Cutting down on repeated lab tests
- Using less reagents and solvents
- Preventing expensive problems when impurities are found too late
- Making chemical pathways better to reduce impurity formation

By finding impurities early in the development process, companies can avoid costly changes to processes or redoing tests later.

LIMITATIONS

Data Quality

AI models rely a lot on having big, clean, and representative sets of data. In the case of predicting impurities, there are common issues with data quality because:

- Limited High-Quality Datasets: AI needs a lot of clean, labeled data.
- But pharmaceutical data is often incomplete, messy, and inconsistent, which makes AI models less accurate.
- Restricted Data Sharing: Much pharmaceutical data is owned by companies.
- Imbalanced Datasets: Some aspects, like certain APIs or impurity types, are well-studied, while others have little data.
- This leads to models that aren’t accurate for less studied areas.
- Lack of Standardized Data Formats: Different labs use different formats for data like HPLC, LC–MS, or NMR.

Low-quality or inconsistent data can cause models to perform poorly, fail to generalize well, and give wrong predictions, which can affect the safety of the product and risk assessments.

Regulatory Acceptance

AI-generated predictions are not yet fully accepted as the main proof by regulatory bodies.

Some challenges include:

- There are no clear rules yet for checking if AI or machine learning models are good enough for impurity analysis.
- Regulators like the FDA, EMA, and CDSCO require models to be explainable, reproducible, and traceable, but many AI models have trouble meeting these needs.
- Deep-learning models can be like "black boxes," making it hard to explain their results during inspections.
- AI predictions are seen as helpful tools, not the main way to decide about impurities or set specifications.

To make AI more widely used, it's important to create standard ways to check models and clearly define what regulators expect.

Model Interpretability

Interpretability means being able to understand how and why a model makes a prediction.In impurity prediction, this is very important because the decisions have big effects on safety and regulations.

Some limitations are:

- Complex models like deep neural networks or graph neural networks are very accurate but not easy to understand.
- Without clear explanations, it's hard to check if the predicted impurities make sense chemically (like if they fit with known reactions).
- Chemical experts might not trust AI predictions if the reasoning behind them is unclear.
- While there are tools like SHAP, LIME, attention mechanisms, and saliency maps, they are still developing and not all standardized for use in pharmaceutical impurity modeling.

Low interpretability makes it harder for regulators to trust AI and slows down its use in quality-by-design and risk assessment processes.

Integration with Existing Pharmaceutical Workflows

- Compatibility issues with legacy systems: LIMS, MES, and ERP systems in pharma are often old and don’t work well with AI tools.
- High implementation cost: Using AI requires spending on cloud computing, GPUs, secure servers, and data systems.
- Skill gap in workforce: Most pharma scientists aren’t trained in AI or machine learning, which slows down adoption and needs a lot of training.
- Resistance to workflow change: Scientists and quality teams may not trust AI predictions and prefer traditional methods.
- Complexity of validation: Pharma processes like GMP, GLP, and QbD need strict documentation; including AI requires new rules and validation steps.
- Maintenance challenges: AI models need constant monitoring, data updates, and recalibration, which adds to the workload.
- Lack of unified AI platforms: Different departments like R&D, quality control, and manufacturing use separate software, making full integration hard.

FUTURE PROSPECTS IN AI-DRIVEN IMPURITY PREDICTION

Use of AI with Automation and Robotics in Quality Control

The use of AI, automation, and robotics is changing how pharmaceutical quality control is done. These tools help speed up and improve the accuracy of checking for impurities. Robotic systems handle sample preparation, extract substances, and load them onto chromatography equipment, which helps cut down on human mistakes. When combined with AI, these systems can automatically look for unusual patterns in test results, tell the difference between types of impurities, and even warn about possible problems before they happen.

AI-powered robots can also quickly test many drug substances and intermediate products.This lets companies assess impurities quickly, which improves how efficiently work gets done, makes results more reliable, and lowers costs. The future might bring self-learning systems that watch how well equipment works in real time and

automatically change settings like temperature, chemical mix, or flow rates to get the best results when checking for impurities.

Key benefits include:

- Fewer mistakes and more consistent results
- Quick decisions with automatic checks on test outcomes
- Better compliance with rules by using digital records of all steps

Digital Twins and Real-Time Impurity Monitorin

Digital twins are virtual copies of real manufacturing processes that update as new data comes in. In drug production, digital twins are becoming important for tracking and predicting impurities.By using detailed models, AI algorithms, process testing tools, and sensors, digital twins can show how impurities form under different conditions. This helps manufacturers make changes early on to stop impurities from building up or catch problems before they affect the final product. Digital twins also support ongoing manufacturing processes and quality planning by helping compare impurity behavior across different sites.

Future uses of digital twins may include:

- Finding impurities as they form during production
- Setting better conditions to reduce impurity levels
- Making real-time changes to processes based on impurity issues

AI-Based Impurity Prediction During Drug Design and Synthesis

AI is becoming more involved in the early stages of creating new drugs, helping to spot possible impurities during the design, development, and selection of chemical processes.

Machine learning models that use large reaction databases, like Reaxys, Pistachio, or CAS, can find reactions that are likely to create harmful by-products, unwanted intermediates, or genotoxic impurities (GTIs).

More advanced models, such as graph neural networks (GNNs), can even predict the exact structure of impurities based on the chemical reactions and molecular structures provided.When adjusting synthetic methods, AI can show how changes in factors like the type of catalyst, solvent used, temperature, or pH may affect the formation of impurities.

This helps chemists pick the best paths to make drugs with fewer impurities. AI can also help estimate how a drug might break down due to light, water, or oxygen during the early design phase, which lowers the risk of problems later on.

Looking ahead, there are several exciting possibilities:

- Using AI to automatically evaluate impurity risks when choosing chemical routes
- Predicting impurities like nitrosamines and other harmful substances with AI
- Combining AI with computer-aided drug design (CADD) to improve the stability of drug molecules
- Fully automated synthesis systems that learn and improve over time to make drugs with minimal impurities

Limitations and Challenges

Despite these advancements, some challenges still exist:

- There's not enough large and high-quality data on impurities to train AI models properly
- There's a need for better data-sharing systems across the pharmaceutical industry
- Regulators are concerned about how clear and reliable AI models are
- Integrating new AI tools with older quality control systems can be difficult
- Setting up robotics and digital twin systems is expensive

CONCLUSION

Artificial intelligence–driven predictive modeling is changing the way the pharmaceutical industry understands and manages API impurities. While traditional analytical techniques like HPLC, LC–MS, GC–MS, and NMR remain essential, they often fall short when dealing with highly complex molecules, massive data volumes, and rising regulatory expectations. AI fills these gaps by offering faster analysis, higher accuracy, and deeper insight into how and why impurities form.Machine learning, deep learning, QSAR models, and reaction-prediction tools allow researchers to foresee impurity formation much earlier in development. These systems can uncover hidden relationships between reaction parameters and impurity outcomes—patterns that are nearly impossible to detect manually. As a result, experimental workload decreases, process optimization becomes more efficient, and decisions can be made more confidently and quickly. Practical applications are already emerging, including AI-based degradation forecasting, retention-time prediction, QSAR-supported mutagenicity assessment, and integration with PAT platforms.

Despite this progress, some challenges still slow down broader adoption. Incomplete or low-quality datasets, difficulty interpreting complex AI models, and ongoing regulatory uncertainties remain significant barriers. Since regulatory bodies emphasize transparency, justification, and model interpretability, the industry must work toward more explainable algorithms and improved data-sharing practices. Without addressing these issues, AI's full impact on impurity prediction will remain limited.In the near future, AI is expected to work hand-in-hand with automation, robotics, digital twins, and smart manufacturing technologies. This shift will transform impurity monitoring into a proactive, real-time, data-driven activity that strongly supports Quality-by-Design (QbD) principles. As models become more accurate, interpretable, and compliant with regulatory expectations, AI will move from being an optional support tool to an essential part of pharmaceutical development and quality assurance

REFERENCES

1. Ahuja, S., & Alsante, K. M. (2017). Handbook of impurities in pharmaceuticals. Elsevier.
2. Balijepalli, S., Hermant, D., Johnson, G. E., & Amberg, A. (2020). Mutagenicity assessment under ICH M7: Use of in silico tools. Regulatory Toxicology and Pharmacology, 118, 104809.
3. Bączek, T., & Kiełbasa, A. (2021). Artificial intelligence in pharmaceutical analysis: A review. Journal of Pharmaceutical and Biomedical Analysis, 192, 113650. https://doi.org/10.1016/j.jpba.2020.113650
4. Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., … Tropsha, A. (2014). QSAR modeling: Where have you been? Where are you going? Journal of Medicinal Chemistry, 57(12), 4977–5010.
5. Dunkel, M., Nagrani, S., & Schmidt, F. (2020). AI-assisted analytical chemistry: Current applications and future perspectives. TrAC Trends in Analytical Chemistry, 130, 115985.
6. European Medicines Agency. (2006). ICH Q3A(R2) & Q3B(R2): Impurities guidelines. EMA.

7. European Medicines Agency. (2006a). ICH Q3A(R2): Impurities in new drug substances. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q3ar2-impurities-new-drug-substances_en.pdf
8. European Medicines Agency. (2006b). ICH Q3B(R2): Impurities in new drug products.https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q3br2-impurities-new-drug-products_en.pdf
9. Food and Drug Administration. (2020). Guidance for industry: Control of impurities in drug substances. U.S. FDA.
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
11. Görög, S. (2018). The importance and challenges of impurity profiling in modern pharmaceutical analysis. Journal of Pharmaceutical and Biomedical Analysis, 161, 110–122.
12. ICH. (2006a). Q3A(R2): Impurities in new drug substances. International Council for Harmonisation.
13. ICH. (2006b). Q3B(R2): Impurities in new drug products. International Council for Harmonisation.
14. ICH. (2016). Q3D: Elemental impurities. International Council for Harmonisation.
15. International Council for Harmonisation. (2021). ICH Q3C(R8): Residual solvents. https://www.ich.org/page/quality-guidelines
16. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. Nature Machine Intelligence, 2(10), 573–584.
17. Kavianpour, B., et al. (2024). AI for chemical and pharmaceutical analysis: Current trends and opportunities. TrAC Trends in Analytical Chemistry, 170.
18. Kulkarni, S. A., Jamadagni, S. N., & Chavan, R. B. (2020). Machine learning approaches in pharmaceutical development. International Journal of Pharmaceutics, 588, 119729.
19. Li, Z., et al. (2023). Deep learning for mass spectrometry data analysis in pharmaceutical quality control. Analytical Chemistry, 95(15), 6221–6230.
20. Machado, B. A. S., Barreto, W. R., da Silva, M. B., & Silva, W. A. (2020). Machine learning in pharmaceutical sciences: Applications in drug development and quality control. Journal of Chemical Information and Modeling, 60(12), 5852–5862. https://doi.org/10.1021/acs.jcim.0c00912
21. Molnar, C. (2022). Interpretable machine learning: A guide for making black-box models explainable. Leanpub.
22. Nayal, A., Singh, R., & Chhabra, D. (2023). AI-enabled prediction of drug degradation pathways: A new direction in stability studies. Journal of Pharmaceutical Sciences, 112(5), 1452–1464.
23. OECD. (2014). Guidance document on the validation of QSAR models. Organisation for Economic Co-operation and Development.
24. Pérez-Neri, I., et al. (2023). Advances in analytical techniques for impurity profiling in pharmaceuticals. TrAC Trends in Analytical Chemistry, 160, 116–122.
25. Ramaswamy, A., Patel, K., & Choi, S. (2022). Integrating AI with process analytical technology for real-time impurity control. International Journal of Pharmaceutics, 617, 121623.
26. Ramesh, M., & Sreshta, S. (2022). Applications of machine learning in predicting pharmaceutical impurities: A systematic review. International Journal of Pharmaceutical Sciences Review and Research, 74(2), 45–52.
27. Reddy, R., Naidu, G., & Kumar, K. (2020). Impurity profiling in pharmaceutical substances. Indian Journal of Pharmaceutical Sciences, 82(1), 1–12.
28. Reker, D., & Schneider, G. (2017). Active-learning strategies in computer-assisted drug discovery. Drug Discovery Today, 22(9), 1407–1415.
29. Schmidt, F., Schneider, P., & Clark, T. (2021). Machine learning-based prediction of side-product formation in organic synthesis. Chemistry–A European Journal, 27(32), 8090–8102.
30. Schwaller, P., Laino, T., & Reymond, J.-L. (2020). Data-driven chemical reaction prediction and its applications. Accounts of Chemical Research, 53(8), 1900–1910.
31. Shah, P., et al. (2022). Machine learning–based prediction of degradation products and impurities in pharmaceuticals. Pharmaceutical Research, 39, 125–138.
32. Singh, D., & Varshney, M. (2020). Applications of machine learning in pharmaceutical impurity profiling. European Journal of Pharmaceutical Sciences, 155, 105558. https://doi.org/10.1016/j.ejps.2020.105558
33. U.S. Food and Drug Administration. (2021). Artificial Intelligence/Machine Learning-Based Software: Action Plan. FDA.
34. Vamathevan, J., et al. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477.
35. Veličković, J., & Dodds, J. (2022). Machine learning approaches for impurity prediction in pharmaceutical development. Computational and Structural Biotechnology Journal, 20, 4354–4365.

36. Xu, W., Yu, K., & Zhang, H. (2023). AI-driven impurity prediction and its regulatory challenges in pharmaceutical development. Trends in Analytical Chemistry, 162, 117004.
37. Yu, L. X., Amidon, G., Khan, M. A., Hoag, S. W., Polli, J., Raju, G. K., & Woodcock, J. (2014). Understanding pharmaceutical quality by design (QbD). AAPS Journal, 16(4), 771–783. https://doi.org/10.1208/s12248-014-9598-3