International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 4 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

INTEGRATING PATIENT GENOMIC PROFILES, MEDICATION RECORDS AND INTERACTION KNOWLEDGE FOR PERSONALIZED MEDICATION SAFETY ANALYTICS - (GENE2DOSE)

[1] Mr. K. Vijay, [2] Sunil Gehlot, [3] Gurijala Manikanta, [4] Amgothu Divya, [5] Guttakindi Kaveri
[1]Assistant Professor, [2345]Students
Department of CSE (Data Science)
Siddhartha Institute of Technology & Sciences, Narapally
vijaykoraveni.cse@siddhartha.co.in, 24tq1a6763@siddhartha.co.in, 24tq1a6752@siddhartha.co.in, 24tq1a6702@siddhartha.co.in, 24tq1a6721@siddhartha.co.in

ABSTRACT

Adverse drug reactions are a leading cause of avoidable hospital admissions, and a considerable share of them can be traced to two factors that are rarely checked together: interactions between the medicines a patient is taking, and genetic variations that change how the patient's body processes a drug. A patient who is a poor metaboliser through the CYP2C19 enzyme, for example, may receive little benefit from clopidogrel, while a patient with certain HLA alleles may develop a severe skin reaction to carbamazepine. This paper presents Gene2Dose, a data integration and analytics system that combines patient genomic profiles, medication records and drug interaction knowledge to provide personalised medication safety analysis.

The system builds a pipeline that ingests pharmacogenomic test results in variant call format, medication orders and dispensing records from the hospital information system, laboratory values such as kidney and liver function, and curated knowledge from public pharmacogenomic guidelines and drug interaction databases. Genetic variants are translated into star-allele diplotypes and metaboliser phenotypes, medicines are mapped to standard drug codes, and all records are linked to a pseudonymised patient identifier in a clinical data warehouse. A knowledge graph connects genes, phenotypes, drugs, enzymes and known interactions so that indirect risks can be traced.

On this integrated data, Gene2Dose performs three kinds of analysis. A rule engine applies guideline recommendations to each patient's phenotype and current medicines, flagging drugs that should be avoided or given at an adjusted dose. A drug interaction module checks every pair of active medicines, including interactions that act through a shared metabolising enzyme whose activity is already reduced by the patient's genotype. A gradient boosting model, trained on historical records, estimates the probability that a patient will experience an adverse drug event in the following thirty days, using the number of flagged risks, age, renal function and polypharmacy as features.

I. INTRODUCTION

Medicines are prescribed on the assumption that a standard dose will produce a similar effect in most patients, yet clinical experience shows wide variation in response. Some patients recover well, others see no benefit, and a few suffer serious harm from the same drug at the same dose. Part of this variation comes from drug interactions, where one medicine alters the absorption, metabolism or effect of another. Another important part comes from inherited differences in the genes that code for drug-metabolising enzymes, transporters and immune system molecules. Older adults and patients with chronic disease, who often take many medicines at once, are especially exposed to both kinds of risk.

Pharmacogenomics, the study of how genes affect drug response, has produced well-established guidance for a growing list of gene and drug pairs. Variants in genes such as CYP2C19, CYP2D6, CYP2C9, VKORC1, TPMT, SLCO1B1 and HLA-B are known to affect the safety or effectiveness of commonly used medicines, including antiplatelet agents, anticoagulants, antidepressants, statins, anti-epileptics and certain cancer drugs. Genetic testing has become cheaper, and some hospitals now order pharmacogenomic panels. However, the test report is usually stored as a separate document, and prescribers must remember to consult it and interpret it whenever a new medicine is started.

Data science and data engineering offer a way to make this information usable at the right moment. A pipeline can translate raw genetic results into standard phenotypes, link them with the patient's medication history and laboratory values, and check them automatically against curated guidelines and interaction knowledge. Machine learning can combine these signals with other risk factors to estimate which patients are most likely to experience harm, so that pharmacists can prioritise their reviews. The challenge lies in integrating heterogeneous data correctly and presenting the results clearly enough for busy clinicians to act on.

II. LITERATURE SURVEY

Research on drug-drug interaction detection has a long history in clinical informatics. Early decision support systems relied on manually curated interaction tables embedded in prescribing software, which checked each new order against the patient's active medicines and displayed warnings. Evaluations of these systems found that they prevented some errors but produced so many low-value alerts that clinicians frequently overrode them, a problem widely described as alert fatigue. Later work refined the knowledge bases by grading interactions according to severity and documentation quality, and by suppressing alerts that are rarely clinically significant. Studies have shown that tiered alerts, in which only severe interactions interrupt the prescriber, improve acceptance considerably. Researchers also note that most interaction checkers treat all patients alike and do not consider individual factors such as kidney function or genotype, which limits their precision.

Pharmacogenomic implementation studies have examined how genetic information can be brought into routine care. Guidelines from international pharmacogenetics consortia translate genotypes into metaboliser phenotypes and give specific therapeutic recommendations for gene and drug pairs, and these guidelines have become the reference for most implementation projects. Pre-emptive testing programmes, in which a panel of genes is tested once and the results are stored for

future use, have been reported by several academic hospitals. Their experience shows that storing results as structured data rather than scanned reports is essential for automated alerts, and that a large proportion of tested patients carry at least one actionable variant. Implementation studies also highlight practical barriers such as limited clinician training, inconsistent nomenclature for alleles and the lack of standard data formats across laboratories.

An important line of research deals with drug-drug-gene interactions, also called phenoconversion. A patient with a normal genotype for a metabolising enzyme may behave like a poor metaboliser if they also take a drug that strongly inhibits that enzyme, and a patient who is already an intermediate metaboliser may be pushed further by a moderate inhibitor. Studies of this effect report that ignoring concomitant medication leads to underestimating the number of patients with functionally impaired metabolism. Computational approaches have been proposed that adjust the predicted phenotype using the inhibitory strength of co-prescribed drugs. This work demonstrates that genetic and interaction checks should not be carried out separately, since the combined effect can be much larger than either one alone.

Knowledge graphs have been applied increasingly in biomedical informatics to represent relationships among drugs, genes, proteins, diseases and adverse effects. By linking public resources on drug targets, metabolic pathways and reported side effects, researchers have built graphs that support queries such as which enzymes metabolise a drug and which other drugs inhibit those enzymes. Graph embedding and link prediction methods have been used to predict previously unreported interactions and adverse effects. Reviews of this work point out that graph-based reasoning helps explain why a risk exists, which is important for clinical acceptance, but that predicted links must be treated carefully and clearly separated from established knowledge when used for patient care.

Machine learning models for adverse drug event prediction have been developed using electronic health record data. Features commonly used include age, sex, number of medicines, specific high-risk drug classes, kidney and liver function, previous admissions and comorbidity indices. Logistic regression, random forests and gradient boosting machines have been compared, and several studies report that ensemble methods give the best discrimination while remaining reasonably interpretable when combined with feature attribution techniques. Very few of these models include pharmacogenomic information, largely because genotype data has been available only for small groups of patients. The literature therefore shows a gap between rule-based pharmacogenomic alerts and data-driven risk prediction, which Gene2Dose addresses by combining guideline rules, interaction checks and a learned risk model on one integrated data platform.

III. SYSTEM ANALYSIS

The system is analysed as a clinical data integration problem followed by knowledge-based and predictive analysis. Four categories of input were identified. Genomic data arrives from the testing laboratory as variant call format files or structured reports listing genotypes for a panel of pharmacogenes. Medication data comes from prescribing, dispensing and administration records in the hospital information system, often with local drug names and free-text dosing. Clinical data includes age, weight,

diagnoses and laboratory values such as creatinine clearance and liver enzymes. Knowledge data comes from published pharmacogenomic guidelines, drug label annotations and drug interaction resources. The analysis showed that each source uses different identifiers and vocabularies, so mappings to standard codes for drugs, genes and laboratory tests are required. Patient privacy demands pseudonymisation and role-based access. The functional requirements are phenotype translation, guideline checking, interaction and phenoconversion checking, risk prediction and a clear display of alerts and recommendations, with every alert traceable to its supporting evidence.

Existing System

In current practice, pharmacogenomic results and medication safety checks are handled separately. When a genetic test is ordered, the laboratory sends a report, often as a PDF attached to the patient record, listing genotypes and a brief interpretation. Prescribers must find and read this report themselves each time they start a new medicine, and many are unaware that it exists. Drug interaction checking, where available, is performed by the prescribing software against a generic interaction table that does not consider the patient's genotype, kidney function or age. Pharmacists review prescriptions manually and rely on reference books and personal knowledge. Adverse drug events are reported voluntarily and analysed rarely. As a result, a patient who is a poor metaboliser may receive a standard dose of a drug that accumulates to toxic levels, and the combined effect of a genetic variant and an inhibiting co-medication is almost never detected.

Disadvantages of Existing System

- Genetic test results are stored as documents that prescribers seldom consult.
- Interaction checks ignore the patient's genotype and metaboliser status.
- Combined drug-drug-gene effects such as phenoconversion go undetected.
- Generic interaction alerts cause alert fatigue and are frequently overridden.
- Local drug names and free-text doses prevent automated checking.
- Kidney and liver function are not considered together with genetic risk.
- Pharmacist reviews depend on memory and printed references.
- There is no estimate of which patients face the highest risk of harm.
- Adverse drug events are under-reported and not analysed systematically.
- Guideline updates are not reflected quickly in clinical practice.

Proposed System

The proposed system, Gene2Dose, is a data integration and analytics platform that brings genomic, medication, clinical and knowledge data together for personalised medication safety. An ingestion pipeline receives variant files and laboratory reports, extracts medication orders and dispensing records from the hospital information system through HL7 FHIR interfaces, and loads laboratory values and demographics. Genotypes are translated into star-allele diplotypes and metaboliser phenotypes, drug names are mapped to RxNorm and ATC codes, and all data is linked under a pseudonymised identifier in a clinical data warehouse. A knowledge graph built in Neo4j links genes, phenotypes, drugs, enzymes and interactions from published guidelines and interaction resources. The analysis layer applies guideline rules for

gene-drug pairs, checks all active drug pairs, adjusts phenotypes for phenoconversion caused by inhibitors and computes a thirty-day adverse drug event risk with a LightGBM model. A dashboard shows each patient's profile, graded alerts with evidence, suggested alternatives or doses and the risk score, and records pharmacist decisions for audit and improvement.

Advantages of Proposed System

- Links genotype, medication, laboratory and demographic data for each patient.
- Translates raw genetic variants into clear metaboliser phenotypes automatically.
- Applies published pharmacogenomic guidelines to every active medicine.
- Detects drug-drug-gene interactions and phenoconversion, not just drug pairs.
- Grades alerts by severity and evidence to reduce alert fatigue.
- Suggests alternative drugs or adjusted doses along with each alert.
- Estimates each patient's thirty-day adverse drug event risk for prioritised review.
- Explains every alert through traceable paths in the knowledge graph.
- Protects patient privacy through pseudonymisation and role-based access.
- Updates rules easily when guidelines or interaction knowledge change.

IV. METHODOLOGY

The methodology begins with data acquisition, in which variant call format files and structured laboratory reports are loaded through an ingestion service, and medication orders, dispensing events, laboratory results and demographics are extracted daily from the hospital information system in FHIR format. Patient identifiers are replaced with pseudonyms using a keyed hash before any processing. Genotypes for the panel genes are called into star-allele diplotypes using a pharmacogene allele caller, and diplotypes are translated into phenotypes and activity scores using guideline tables. Drug names and doses are normalised to RxNorm and ATC codes through dictionary matching and fuzzy string matching, with unresolved entries sent to a pharmacist for mapping. Guideline recommendations, enzyme substrates, inhibitors and inducers, and interaction pairs are loaded into the knowledge graph. Rule checks and graph queries generate alerts, phenotypes are adjusted for strong and moderate inhibitors, and features such as alert counts, medicine count, age and creatinine clearance are used to train the LightGBM risk model with stratified cross validation and SHAP explanations.

SYSTEM ARCHITECTURE

The architecture of Gene2Dose separates a data integration track and a knowledge track that meet in the analysis layer. The Data Sources layer contains the genomics laboratory, the hospital information system with prescribing and dispensing records, the laboratory information system and the published guideline and interaction resources. The Ingestion and Privacy layer includes the variant file loader, FHIR extractors and the pseudonymisation service. The Transformation layer performs allele calling and phenotype translation, drug code normalisation and laboratory value harmonisation. The Storage layer consists of the clinical data warehouse in PostgreSQL for patient-level records and the Neo4j knowledge graph for genes, drugs,

enzymes and interactions. The Analysis layer contains the guideline rule engine, the interaction and phenoconversion checker and the LightGBM adverse event risk model. The Presentation layer provides the medication safety dashboard for pharmacists, patient-specific alerts for prescribers and an audit log that records accepted and overridden alerts for continuous improvement.

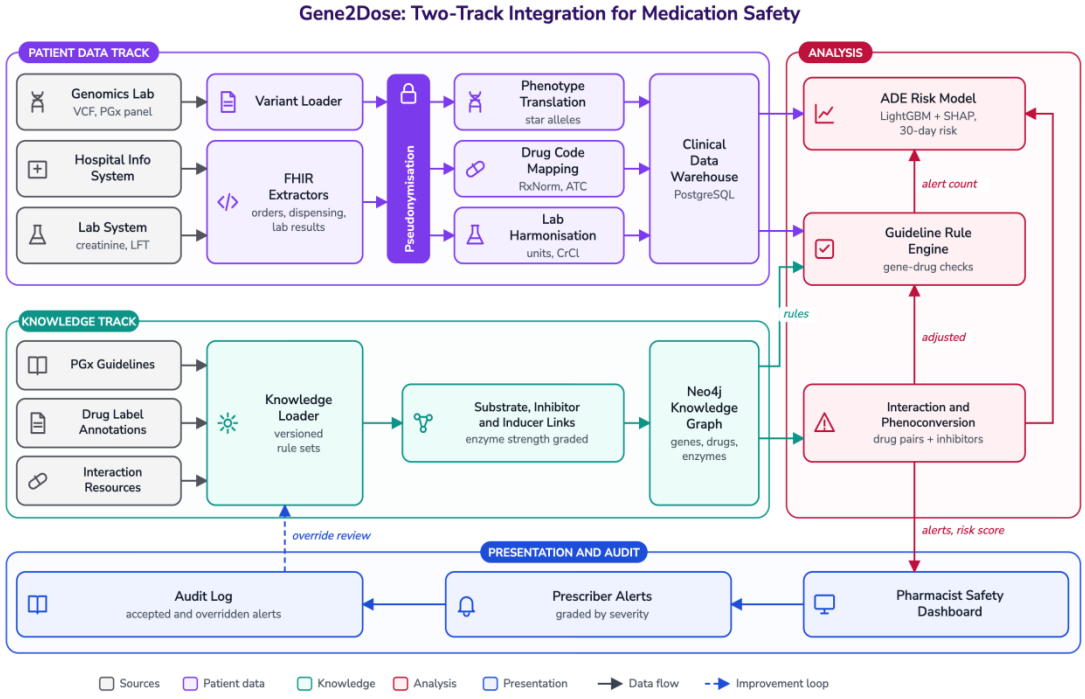


Fig. 1 System Architecture

V. RESULT AND OUTPUT

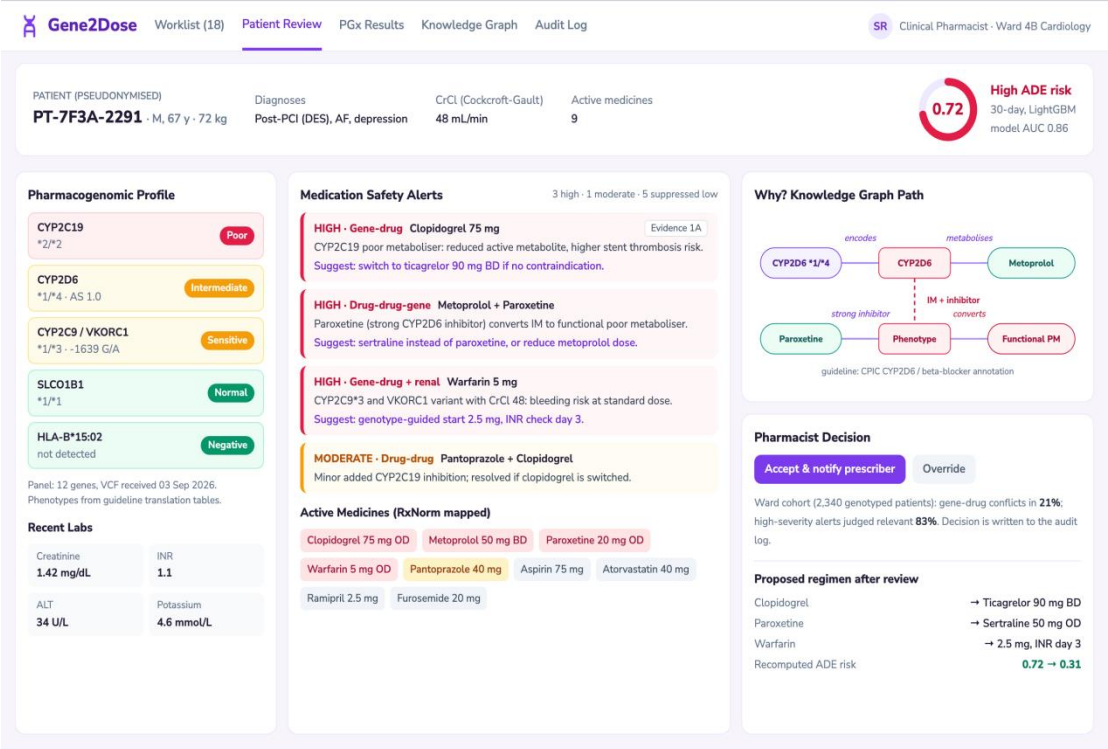


Fig. 2 Result and Output Screen

VI. CONCLUSION

This paper presented Gene2Dose, a system that integrates patient genomic profiles, medication records, laboratory data and drug interaction knowledge to provide personalised medication safety analytics. By translating genetic results into phenotypes and linking them with coded medication data in a common warehouse and knowledge graph, the system makes pharmacogenomic information available at the time when prescribing decisions are reviewed, rather than leaving it in a report that is seldom read.

The retrospective evaluation on 2,340 de-identified patient records showed that such integration uncovers risks that conventional checks miss. Gene-drug conflicts were found in 21 percent of patients, and a noticeable number of additional risks arose only when genotype and co-medication were considered together through phenoconversion. The LightGBM risk model reached an area under the ROC curve of 0.86, helping pharmacists focus their reviews on patients most likely to experience an adverse drug event, and 83 percent of high-severity alerts were judged clinically relevant.

The work also shows the value of careful data engineering in clinical settings. Mapping local drug names to standard codes, harmonising laboratory units and consistently translating alleles were essential steps, and errors in any of them would produce misleading alerts. Grading alerts by severity and evidence and showing the knowledge graph path behind each one helped keep the number of interruptions manageable and increased the trust of pharmacists in the output.

Future work will expand the gene panel, integrate the alerts directly into electronic prescribing so that they appear at the moment of ordering, and carry out a prospective

study in hospital wards to measure the effect on adverse drug events and length of stay. Incorporating dose-response models for drugs with narrow therapeutic windows and supporting population-specific allele frequencies relevant to Indian patients are further directions for development.

REFERENCES

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu, bhargavi nalacheruve. (2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.