QUANTUM INTELDRUG - QUANTUM COMPUTING AND MACHINE LEARNING FOR ENHANCED CANCER DRUG DISCOVERY

Project Reference No.: *49S_BE_2665*

College : Adichunchanagiri Institute of Technology, Chikkamagaluru
Branch : Department of Computer Science and Engineering (Data Science)
Guides : Prof. Shalini I S
Dr. Adarsh M J
Students : Ms. M M Bhavya
Ms. Meghana N S
Ms. Rishitha N

Keywords:

Quantum Computing, Machine Learning, Cancer Drug Discovery, Supper Vector Machne, Random Forest, Quantum Support Vector Machine (QSVM) and Morgan Fingerprints.

Introduction:

Drug discovery is one of the most important and challenging areas in pharmaceutical and biomedical research, as it involves identifying molecules that can effectively treat diseases. Traditional drug discovery methods are highly time consuming, expensive, and often take many years with a low success rate. To overcome these limitations, The Quantum IntelDrug - Quantum Computing and Machine Learning for Enhanced Cancer Drug Discovery project focuses on predicting the anticancer activity of chemical compounds using advanced computational techniques. The system accepts molecular structures in the form of SMILES strings, which are converted into morgan fingerprints using RDKit for feature extraction.

These features are analyzed using three predictive models: Support Vector Machine (SVM), Random Forest, and Quantum Support Vector Machine (QSVM). This allows comparison between machine learning methods and quantum inspired approaches for molecular activity prediction.

The project is highly significant in the field of cancer research, where accurate prediction of active compounds can reduce the need for extensive laboratory screening and speed up drug development.

A Streamlit-based web interface is developed to enable users to input molecular data, visualize structures, choose models, and obtain predictions with confidence scores.

By integrating machine learning, quantum computing, and cheminformatics, the project provides an innovative and practical framework for early-stage cancer drug discovery.

Objectives:

1. To develop a hybrid prediction system combining Machine Learning and Quantum Computing algorithms for anticancer drug activity prediction.
2. To process SMILES-based molecular data and convert it into morgan fingerprints using RDKit for effective feature extraction and analysis.
3. To classify compounds as active or inactive and compare the performance of machine learning and quantum models based on accuracy and confidence scores.
4. To build a user friendly Streamlit web application for molecular visualization, prediction reporting, and result export in CSV/PDF formats.

Methodology:

The project focus on the integration of Quantum Computing and Machine Learning techniques to enhance the process of cancer drug discovery. The work is be carried out in a systematic sequence of steps to ensure accurate prediction and efficient analysis of drug molecules.

1. Initially, the required dataset related to cancer drug compounds are collected from reliable biomedical databases such as ChEMBL and other publicly available molecular datasets. The dataset consist of molecular structures, biological activity values, and compound properties relevant to cancer treatment.
2. In the preprocessing stage, the collected data is cleansed by removing missing values, duplicate records, and irrelevant attributes. The molecular structures available in SMILES format are then converted into machine-readable numerical representations. For feature extraction, Morgan fingerprints are generated using computational chemistry tools such as RDKit. These fingerprints helps in representing the chemical characteristics of drug compounds.
3. The processed dataset is divided into training and testing sets for model development and evaluation. Machine learning algorithms such as Support Vector Machine (SVM) and Random Forest are first implemented to predict whether a compound is active or inactive against cancer molecules.
4. Further, a Quantum Support Vector Machine (QSVM) model need to be developed, where quantum circuit is used to encode molecular features into quantum states.
5. The performance of classical and quantum models are evaluated using accuracy, confidence score, and comparison metrics.

Finally, the results obtained from both approaches are compared to analyze the effectiveness of Quantum Computing and Machine Learning in improving prediction accuracy for cancer drug discovery.

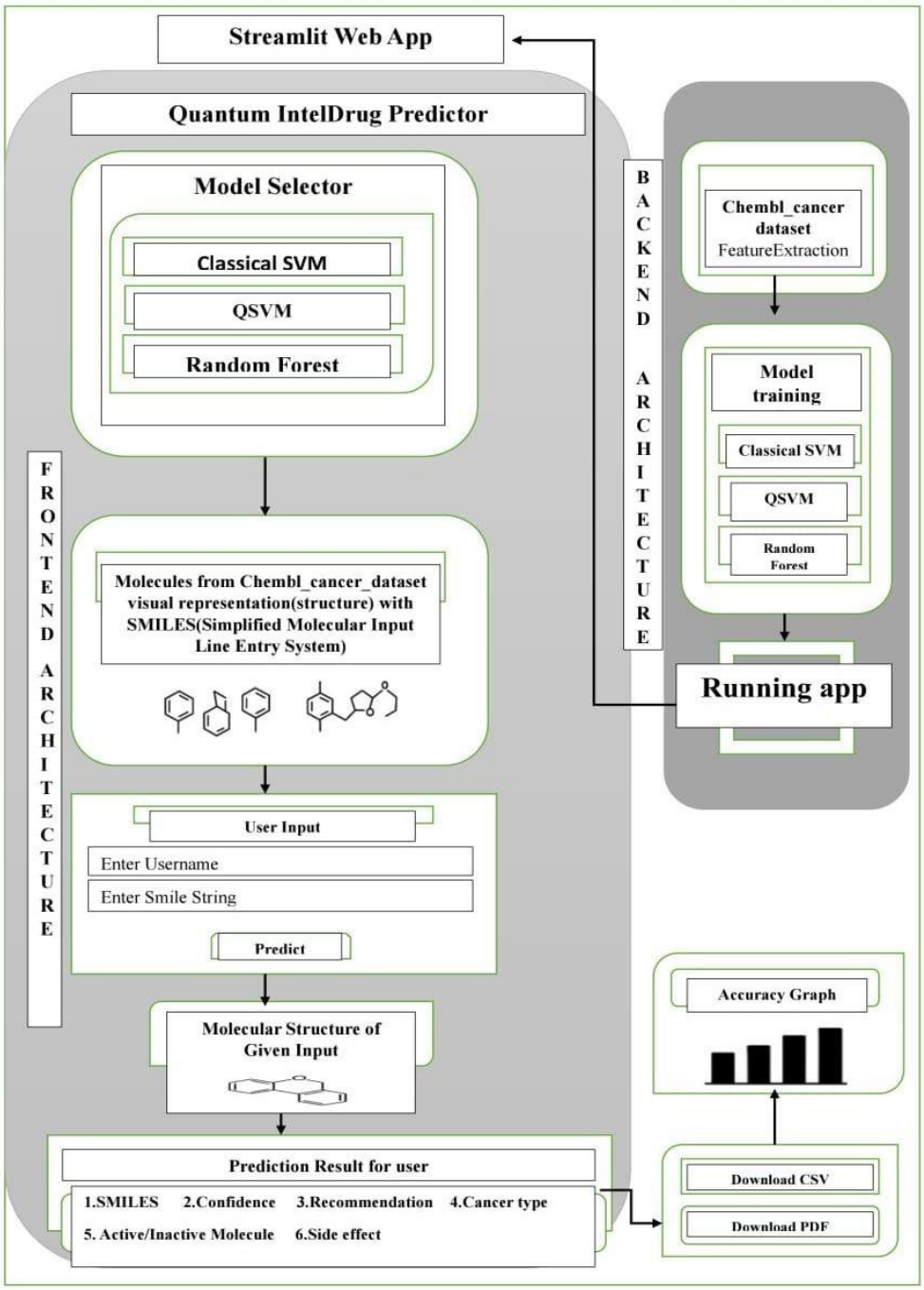

Figure 1: System Architecture

Result and Conclusion:

The proposed system predicts the anticancer activity of chemical compounds using both machine learning and quantum computing approaches. Molecular data in the form of SMILES strings are processed and converted into morgan fingerprints using RDKit, enabling effective feature extraction and representation.

The models, namely Support Vector Machine (SVM) and Random Forest, generates reliable classification results for identifying active and inactive compounds. The Quantum Support Vector Machine (QSVM) model captures complex molecular relationships using quantum feature mapping and entanglement concepts, showing promising predictive capability. Comparative analysis of the model indicate that these

approaches perform efficiently on structured datasets, while quantum-based methods provide advantages in handling high-dimensional and complex feature spaces. The system also generates confidence scores, predicted activity class, possible side effects, and associated cancer types, providing a comprehensive understanding of each compound.

A Streamlit based web application is developed to enable user interaction, molecular visualization, and real-time prediction. Graphs and charts are also be used to compare model accuracy and performance.

In conclusion, the integration of Quantum Computing and Machine Learning enhances the efficiency of drug discovery by improving prediction accuracy, reducing experimental cost, and accelerating early-stage screening of anticancer compounds.

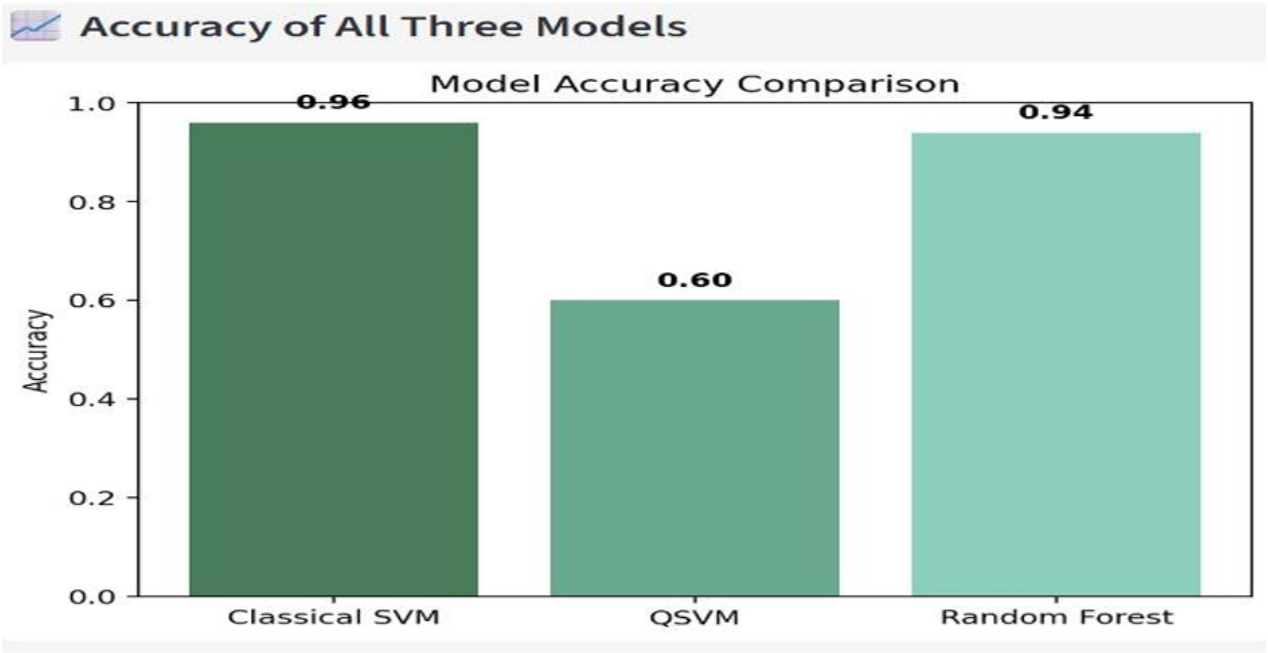

Figure 2: Comparative Accuracy Analysis of SVM, Random Forest, and QSVM Models

Project Outcome and Industry Relevance:

1. The project resulted in the development of an intelligent and user-interactive system for predicting anticancer drug activity using advanced computational techniques. It integrates machine learning, quantum computing, and cheminformatics into a unified framework, providing an efficient platform for molecular analysis and prediction.
2. The system is capable of identifying potential drug candidates at an early stage, thereby reducing the dependency on time-consuming and expensive laboratory experiments.
3. In the pharmaceutical and biotechnology industries, such a system can significantly accelerate the drug discovery process and improve decision-making by focusing only on promising compounds. The use of Quantum Support Vector Machine (QSVM) highlights the growing relevance of quantum computing in solving complex biomedical problems. The developed web application enhances accessibility for

researchers, enabling easy input, visualization, and report generation in CSV and PDF formats.

4. Overall, the project aligns with current industry trends such as AI-driven drug discovery and precision medicine, making it highly relevant for real-world applications and future technological advancements.

Working Model vs. Simulation/Study:

The project is primarily a simulation-based study combined with a functional software implementation. It does not involve a physical working model but focuses on computational techniques for drug discovery using machine learning and quantum computing concepts. The system simulates the behavior of classical machine learning and quantum machine learning models to analyze and predict anticancer activity of chemical compounds.

A working prototype has been successfully developed in the form of a Streamlit web application, which provides an interactive platform for users to input molecular data, visualize chemical structures, and obtain prediction results. The system integrates models such as Support Vector Machine (SVM) and Random Forest, along with quantum-based models like Quantum Support Vector Machine (QSVM). Hence, the project can be classified as a software based working model supported by simulation of quantum algorithms, demonstrating practical implementation and real-time usability.

Project Outcomes and Learnings:

Project Outcomes include:

1.Predictive Tool:

The primary practical outcomes are functional, user interactive Streamlit based web platform that predicts a molecule's anticancer potential (Active or Inactive).

2.Comprehensive Output:

Users receive not just only the classification, but also a confidence score, an estimation of potential side effects, and the associated cancer types that the compound might treat.

3.Molecular Analysis:

The Quantum IntelDrug Predictor System provides the capability to visualize molecular structures from a simple SMILES string input, making abstract computational results more accessible for experimental researchers.

4.Comparative Analysis:

The project enables a direct comparison between the performance of classical machine learning models (SVM and Random Forest) and the Quantum Support Vector Machine (QSVM). This will be crucial for evaluating the potential quantum advantage in a real-world cheminformatics task.

Learnings from the Project through the process of designing, implementing, and analyzing this project, several important learnings are achieved:

1. Practical experience gained by integrating machine learning and quantum computing techniques for real world applications.
2. Skills are developed in handling chemical datasets, including processing SMILES strings and generating molecular fingerprints.
3. Knowledge acquired in building and deploying interactive web applications using Streamlit, improving user accessibility and experience.
4. The importance of model evaluation and comparison are understood, especially when exploring emerging technologies like quantum machine learning.
5. Understanding of drug discovery concepts, including anticancer prediction, toxicity, and side effect analysis, will be enhanced.
6. Problem solving and system design abilities strengthened by working on an end-to-end pipeline, from data preprocessing to result visualization and reporting.

References:

1. Li, J., Alam, M., Sha, C. M., Wang, J., Dokholyan, N. V., & Ghosh, S. 2021). “Drug discovery approaches using quantum machine learning.” Proceedings of the 24th International Symposium on Quality Electronic Design (ISQED), IEEE, Volume 2021, Pages 173–178. https://arxiv.org/abs/2104.0074p6
2. 2. Avramouli, M., Savvas, I. K., Vasilaki, A., & Garani, G (2023) “Unlocking the potential of quantum machine learning to drug discovery” Preprints,2023050473. https://doi.org/10.20944/preprints202305.0473.v1
3. Shara, J. (2024). “An Overview of Quantum Machine Learning and Drug Discovery”. International Journal of Advances in Engineering And Management (IJAEM), 6(08), 174– 183. https://www.researchgate.net/publication/383122139
4. Deepa Priya, V., Gnana Jersha, J., Madhubhavani, G., Mangalya, T., Balaharish Alias Yogesh, N., & Chellapandi, T. R. (2025). “Quantum computing in drug discovery”. International Research Journal on Advanced Engineering and Management (IRJAEM), 3(3), https://doi.org/10.47392/IRJAEM.2025.0123
5. Bertl, M., Mott, A., Sinno, S., & Bhalgamiya, B. (2025). “Quantum machine learning in precision medicine and drug discovery – A game changer for tailored treatments?”arXiv preprint arXiv 2502.18639. https://arxiv.org/abs/2502.18639

Future Scope:

The future scope of this project includes:

1. ADMET prediction will be integrated to improve early evaluation of drug safety and effectiveness.
2. The system will be connected with real-time clinical data sources (EHRs and clinical trial databases) for continuous learning and improved accuracy.

3. A Quantum-AI platform will be developed to enable collaboration among researchers, startups, and pharmaceutical companies.
4. The scalability and efficiency of quantum machine learning models will be enhanced, and they will be combined with deep learning techniques for better performance.
5. The project will be expanded to support multiple diseases, and it will integrate drug databases and automated reporting systems for improved usability.