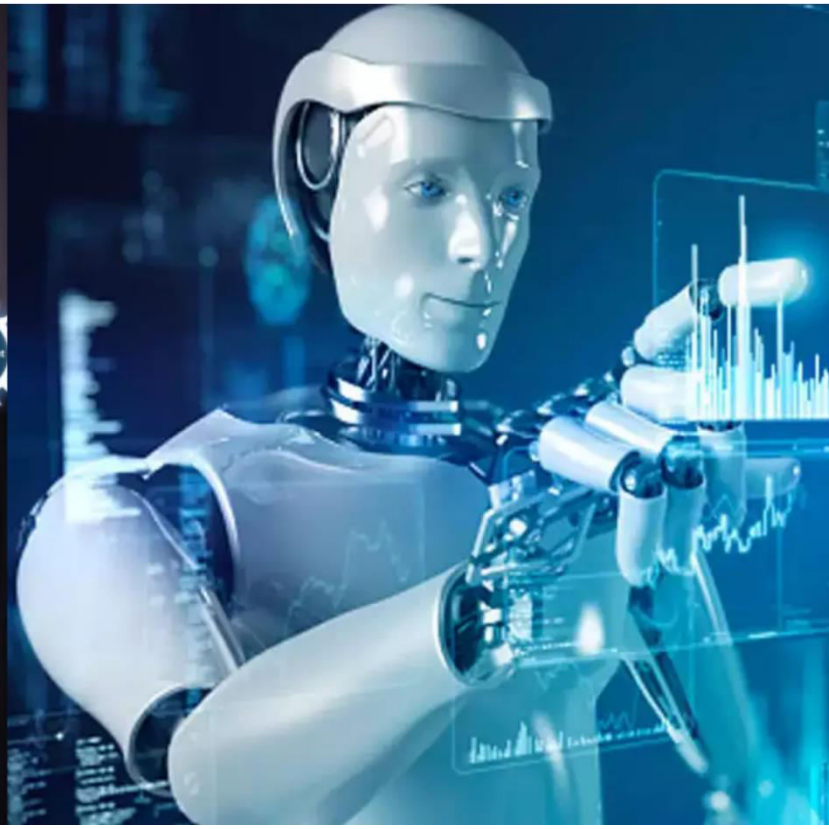
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

AI Enhanced Drug Prediction and Testing for Faster Development in Breast Cancer

Rakshitha C R[1], Sahana M G[2], Sahana S[3], Sowjanya C[4], Rajani D[5]

UG Student, Department of Computer Science and Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India[1 2 3 4]

Assistant Professor, Department of Computer Science and Engineering, GSSS Institute of Engineering and Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India[5]

ABSTRACT: Breast cancer still remains one of the major health issues worldwide; its effective treatment is often hampered by the traditional drug development pipeline. Laboratory-based screening of chemical compounds requires extensive biological testing at high research cost with long validation cycles, eventually leading to slow progress and high attrition rates. Artificial intelligence rapidly analyses molecular structures and can predict properties that influence drug safety and efficacy. This paper presents the development of an AI-driven computational framework that supports early-stage breast cancer drug discovery. The system takes molecular inputs in the form of SMILES, converts them into fingerprints using RDKit, and passes them for evaluation through two predictive models, namely, toxicity classification and DTI prediction against breast cancer important protein targets such as ESR1, ERBB2, EGFR, and PGR. Furthermore, a BRICS-based molecular design module will generate new chemical structures for further screening. A Stream lit web interface exposes the system to users in a highly usable form and allows single predictions, batch screening, as well as visualization.

KEYWORDS: Breast cancer, Drug discovery, Toxicity prediction, Drug–Target Interaction, Molecular fingerprinting, BRICS algorithm, RDKit, Machine learning.

I. INTRODUCTION

Among the most prominent reasons for cancer-related deaths in women is breast cancer, which is a significant priority for researchers seeking effective leads for a cure. In short, the conventional way to develop a drug is a process of chemical library synthesis, wet lab experiments, and a series of biological validations. This process is extremely resource-intensive, extremely time-consuming, but is generally a low success rate process. The most significant setback that is usually experienced is that potential leads that are discovered are apt to fail later on, because of toxicity, insufficient protein binding, and pharmacokinetics, which are hard to predict in the initial stages of a manual process.

The rising availability of computational power is making AI a revolutionizing force in the drug discovery industry. By learning from chemical and biological databases, AI models are learning structure-property correlations, providing instant predictions that are helpful in research work. Toxicity prediction, approximation of drug-target interactions, and the design of molecular variations with the use of AI can cut the cost of experiments by an order of magnitude.

This article discusses the development of a single, AI-assisted tool that can facilitate discoveries in the drug development of breast cancer in the initial stages. The tool is a mixture of toxicity prediction, protein binding prediction, BRICS molecule generation, and a web interface for user interactions. Such a tool with diverse functionalities is guaranteed to bring the timelines of discoveries closer to the researchers, even when they are non-professionals, unlike at present.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

II. LITERATURE SURVEY

Current areas of application being increasingly explored by researchers involve the use of AI algorithms in drug development, cancer biology, and chemical property prediction. Various studies have pointed out that AI accelerates discovery cycles and improves candidate selection.

Herráiz-Gilet et al. (2025) identified that AI models are able to identify relationships in structure-biology space, enabling the efficient repurposing of drugs and optimization of cancer therapies. Gandhi and Kumar (2024) have discussed how AI improves diagnosis and treatment planning in breast cancer based on clinical data analysis, imaging, and molecular pathways [1].

Dhudum et al. (2024) showed that deep learning and graph-based models significantly improve compound screening and toxicity evaluation. Adamopoulos et al. (2024) further reported integration of machine learning with biological datasets, improving the prediction accuracy for anti-cancer compounds [3].

Blanco-González (2023) highlighted opportunities and challenges when adopting AI in pharmaceutical research: dataset limitations, regulatory constraints, and system interpretability. Wang et al. (2023) presented a review about a decade of anti-cancer drug design driven by AI and emphasized the growth of generative models for molecule synthesis [6].

Askr 2022 discussed the role of deep learning in the prediction of molecular behaviour and drug sensitivity. On the other hand, Chiu et al., 2019 combined neural networks with genomic profiles in predicting tumour drug response [10].

Taken cumulatively, these findings suggest that AI can drastically reduce time in discovery, analyse intricate molecular spaces, and offer better early-stage prediction accuracy. However, certain challenges pertaining to data quality, model transparency, and interpretability limit further application—thus motivating the need for integrated, user-friendly tools such as the system proposed herein.

III. METHODOLOGY

We present a modular computational pipeline that systematically ensures reproducibility of each step, from data pre-processing to prediction.

A. Molecular Input and Pre-processing

Chemical structures are in SMILES format. Each input is checked by RDKit, structural inconsistencies are removed, aromaticity is standardized and the molecules are turned into canonical SMILES. Invalid structures are automatically marked as such.

B. Feature Extraction

RDKit is used to calculate the:

- Morgan fingerprints
- Physicochemical descriptors: logP, molecular weight, H-bond donors/acceptors, etc.
- structural keys and subgraph features

These features form the input vectors for the machine learning models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

C. Toxicity Prediction Model

A Random Forest classifier will be trained on curated toxicology datasets (e.g., Tox21) to predict whether a molecule is toxic or non-toxic. The model outputs:

- Class label
- Toxicity probability score
- Feature importance insights

D. Drug–Target Interaction Prediction

A multi-class classifier that estimates the interaction probabilities with four key breast cancer proteins:

- ESR1
- ERBB2
- EGFR Abbreviation of epidermal growth factor receptor.
- PGR

The model uses fingerprint vectors in conjunction with protein-specific training data.

E. Novel Molecule Generation BRICS Algorithm

BRICS is a fragmentation technique that decomposes molecules into fragments, which follow the chemically meaningful boundaries, and subsequently recombines these fragments to form a new molecule, obeying some drug-likeness constraints.

The generated molecules then flow back into the toxicity and DTI pipelines for evaluation.

F. Stream lit-Based User Interface

The interface supports the following:

- Single SMILES prediction
- Batch Molecule Screening
- BRIC molecular synthesis
- Toxicity and DTI visualization
- downloadable output reports

Caching provides for seamless performance even for large-scale predictions.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

IV. EXPERIMENTAL RESULTS

Figure 1: Toxicity Predictor-SMILES Lookup

This interface illustrates the workflow of toxicity prediction of the proposed system. A known drug name is entered by the user, and the application populates its corresponding SMILES representation through a live lookup feature. When the SMILES string is populated, the Random Forest toxicity model assesses the structure for a probability score of toxicity. In the case of Docetaxel, the model predicted 0.513, which falls into the moderately toxic range and agreed well with the clinically established toxic profile of this drug. This feature enables rapid validation of existing drugs or the investigation of new compounds without having to manually look up chemical databases.

Figure 2: Toxicity Predictor

This figure shows the ease of use of the toxicity prediction tab by incorporating SMILES lookup, manual input, and model selection. Real-time validation of SMILES and generation of the toxicity score immediately after prediction are built into the interface. This design further improves its usability, particularly for researchers and students without much cheminformatics background. The interface guarantees ease of navigation and speed in accessing predictive tools within the system.

Figure 3: Drug–Target Interaction prediction (DTI)

The DTI interface allows users to estimate the probability of a molecule interacting with major breast cancer targets such as ESR1, ERBB2, EGFR, and PGR. Given a SMILES string and a target protein, this model returns a probability that a given compound is active against a selected protein. The provided example illustrates that the c1ccccc1 (benzene) molecule showed low interaction probability-0.328 toward the ESR1 receptor, which confirmed the non-active structure classification by the model. This module allows for fast prioritization of molecules for further biological evaluation.

Figure 4: Batch Toxicity Scoring Using CSV Input

Batch scoring supports the high-throughput toxicity screening by providing a facility to upload a CSV file containing multiple SMILES entries. The system processes the file and presents the user with a table listing the corresponding toxicity probabilities for each molecule. This facility will prove to be much useful if one is testing large compound libraries from ChEMBL or PubChem datasets. The results can be downloaded for offline analysis, thus making the system suitable for large-scale computational filtering before laboratory testing.

Figure 5: BRICS-based design and candidate ranking

This section introduces the molecular design module that uses BRICS fragmentation and recombination to generate novel compounds. The user has to specify the number of molecules to be generated and the target protein on which the scoring is to be done. The generated structures are filtered for drug-likeness properties, toxicity predictions, and DTI scores. This module enables early ideation in drug discovery with the creation of chemically feasible candidates that could be directly fed into further computational or experimental workflows.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

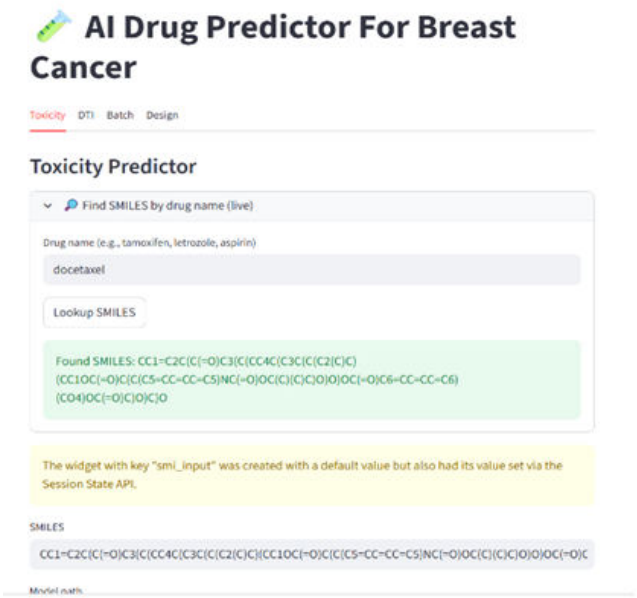


Fig.1: Toxicity Predictor-SMILES Lookup

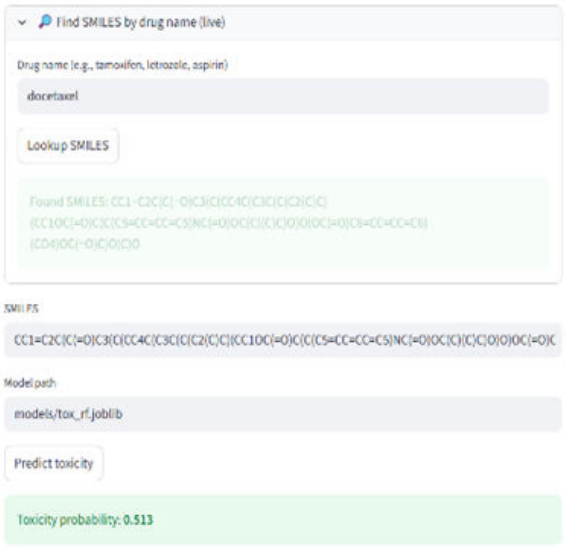


Fig.2: Toxicity Predictor

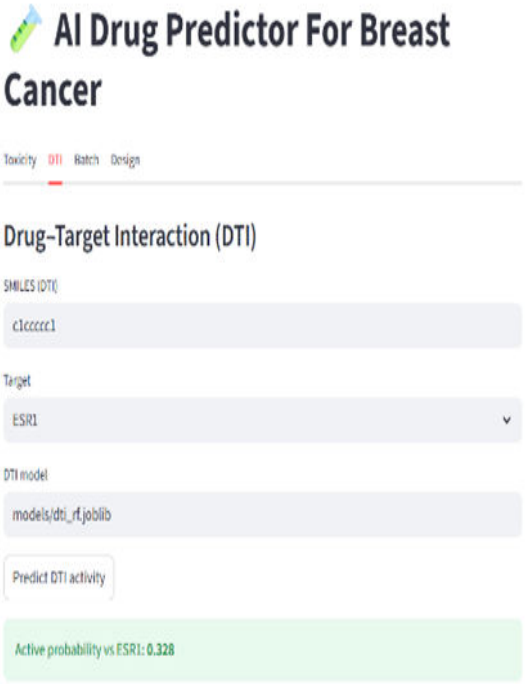


Fig.3: Drug–Target Interaction prediction (DTI)

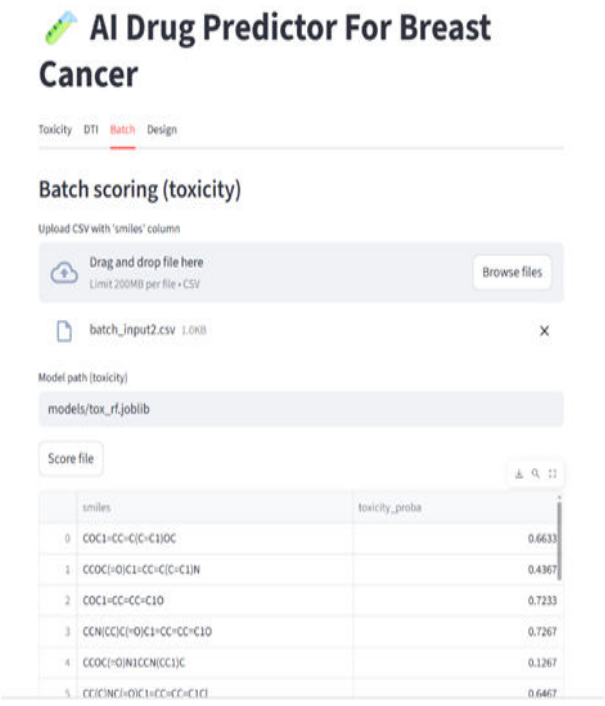


Fig.4: Batch Toxicity Scoring Using CSV Input

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

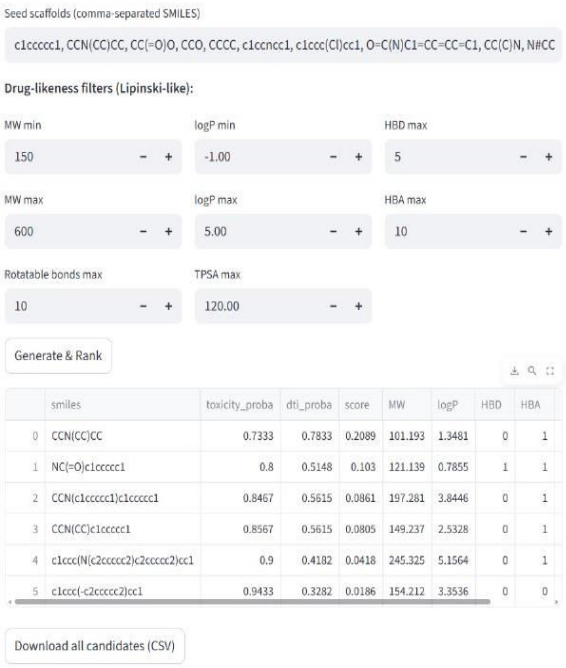


Fig.5: BRICS-based design and candidate ranking

V. CONCLUSION

The system illustrates a way in which AI can contribute to the process of early-stage drug discovery in breast cancer. This framework uses predictive modelling, cheminformatics techniques, and automated molecule design to reduce dependence on laboratory screening to accelerate initial candidate selection. An integrated UI makes the tool available for academic and research use by letting users explore and test molecular structures without any prior programming experience.

Future work may incorporate transformer-based chemical encoders, molecular docking simulations, expanded protein targets, and real-world laboratory collaboration for the validation of top-scoring molecules.

REFERENCES

[1] Herráiz-Gilet, S., López-Santamaría, M., and Biarnés-Pérez, M., "Artificial Intelligence-Based Methods for Drug Repurposing and Cancer Treatment Development," *Computational Biology and Chemistry*, vol. 109, pp. 1–18, 2025.

[2] Adamopoulos, C., Karyofyllis, P., Vrahatis, M.N. et al., "Fusing Artificial Intelligence and Machine Learning for Anti-Cancer Drug Discovery," *Cancers*, vol. 16, no. 1, pp. 1–25, 2024.

[3] Dhudum, R., Ganeshpurkar, A., Pawar, A., "Revolutionizing Drug Discovery: A Comprehensive Review of AI Applications," *Journal of Drug Development Research*, vol. 13, no. 4, pp. 56–78, 2024.

[4] Gandhi, H., Kumar, K., "Artificial Intelligence for the Management of Breast Cancer: An Overview," *Healthcare Analytics*, vol. 8, pp. 1–15, 2024.

[5] Landrum, G., "RDKit: Open-Source Cheminformatics," *RDKit Documentation*, Release 2024.03, 2024. [Online].

[6] Blanco-González, A., "The Role of Artificial Intelligence in Drug Discovery: Challenges, Opportunities, and Strategies," *Drug Discovery Today*, vol. 28, no. 5, pp. 1–12, 2023.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412062|

[7] Wang, L., Yu, J., Hu, K. et al., "Advances of Artificial Intelligence in Anti-Cancer Drug Design: A Review of the Past Decade," *Pharmaceuticals*, vol. 16, no. 2, pp. 1–30, 2023.
[8] Yang, X., Wang, Y., Byrne, R., et al., "A Review of Machine Learning for Drug Discovery," *Journal of Drug Design*, vol. 18, no. 2, pp. 45–67, 2023.
[9] Kim, S., et al., "PubChem 2023 Update: New Data Content and Improved Web Interfaces," *Nucleic Acids Research*, vol. 51, pp. D1373–D1380, 2023.
[10] Askr, H., "Deep Learning in Drug Discovery: An Integrative Review and Future Challenges," *Molecules*, vol. 27, no. 2, pp. 1–20, 2022.
[11] Zhang, Q., Aires-de-Sousa, J., et al., "ADMET Prediction Using Machine Learning Models," *Frontiers in Pharmacology*, vol. 13, pp. 1–12, 2022.
[12] Bajorath, J., "Chemoinformatics Tools for Modern Drug Discovery," *ACS Chemical Reviews*, vol. 121, no. 15, pp. 8736–8758, 2021.
[13] Chen, L., Cruz, A., Ramsey, S. et al., "Machine Learning for Toxicity Prediction," *Journal of Chemical Information and Modeling*, vol. 61, no. 3, pp. 1–12, 2021.
[14] Jiménez-Luna, J., Grisoni, F., Schneider, G., "Drug Discovery with Explainable AI," *Nature Machine Intelligence*, vol. 3, pp. 25–36, 2021.
[15] Irwin, J., Shoichet, B., "ZINC15: Ligand Discovery for Everyone," *Journal of Chemical Information and Modeling*, vol. 60, no. 12, pp. 6065–6073, 2020.
[16] Chiu, Y.C., Chen, Y., Zhang, T. et al., "Predicting Drug Response of Tumors from Integrated Genomic Profiles Using Deep Neural Networks," *BMC Medical Genomics*, vol. 12, no. 1, pp. 1–15, 2019.
[17] Vamathevan, J., Clark, D., et al., "Applications of Machine Learning in Drug Discovery and Development," *Nature Reviews Drug Discovery*, vol. 18, pp. 463–477, 2019.
[18] Brown, N., Ertl, P., et al., "Designing Novel Compounds Using BRICS Fragmentation," *Journal of Computational Chemistry*, vol. 40, pp. 936–949, 2019.
[19] Öztürk, H., Özgür, A., Ozkirimli, E., "DeepDTA: Deep Drug–Target Binding Affinity Prediction," *Bioinformatics*, vol. 34, no. 17, pp. 821–829, 2018.
[20] Wishart, D., et al., "DrugBank 5.0: Expanding Drug Knowledge," *Nucleic Acids Research*, vol. 46, pp. D1074–D1082, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details