International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

DEEPSIDE A DEEP LEARNING FRAMEWORK FOR DRUG SIDE EFFECT PREDICTION

VADDI SRIVALLIDEVI, Sk. Ruhi

Associate professor,
Department of MCA
vsrivallidevi95@gmail.com
B V Raju College, Bhimavaram
(2285351107)
Department of MCA
shaikruhi1626@gmail.com
B V Raju College, Bhimavaram

ABSTRACT

Drug failures due to unforeseen adverse effects at clinical trials pose health risks for the participants and lead to substantial financial losses. Side effect prediction algorithms have the potential to guide the drug design process. LINCS L1000 dataset provides a vast resource of cell line gene expression data perturbed by different drugs and creates a knowledge base for context specific features. The state-of-the-art approach that aims at using context specific information relies on only the high-quality experiments in LINCS L1000 and discards a large portion of the experiments. In this study, our goal is to boost the prediction performance by utilizing this data to its full extent. We experiment with 5 deep learning architectures. We find that a multi-modal architecture produces the best predictive performance among multi-layer perceptron-based architectures when drug chemical structure (CS), and the full set of drug perturbed gene expression profiles (GEX) are used as modalities. Overall, we observe that the CS is more informative than the GEX. A convolutional neural network-based model that uses only SMILES string representation of the drugs achieves the best results and provides 13:0% macro-AUC and 3:1% micro-AUC improvements over the state-of-the-art. We also show that the model is able to predict side effect-drug pairs that are reported in the literature but was missing in the ground truth side effect dataset.

Keywords: deep learning, drug side effect prediction, LINCS L1000 dataset, gene expression profiles, drug chemical structure, convolutional neural network, SMILES string representation.

INTRODUCTION

The development of new pharmaceuticals is a complex and expensive process, often marred by unforeseen adverse effects that can emerge during clinical trials. These side effects not only pose significant health risks to participants but also lead to substantial financial losses for pharmaceutical companies. Predicting these adverse effects early in the drug development process is therefore crucial. The advent of computational methods and the availability of large-scale biological datasets have opened new avenues for improving the prediction of drug side effects. One such promising approach is the use of deep learning frameworks, which can leverage vast amounts of data to uncover patterns and associations that might elude traditional methods. The Library of Integrated Network-based Cellular Signatures (LINCS) L1000 dataset is one such resource, providing extensive cell line gene expression data perturbed by various drugs. This dataset serves as a rich knowledge base for extracting context-specific features relevant to drug responses. The current state-of-the-art methods for side effect prediction using LINCS L1000 data often focus on high-quality experiments, discarding a significant portion of the available data. While this approach aims to maintain high data quality, it potentially overlooks valuable information present in the discarded experiments. In our study, we

aim to enhance prediction performance by utilizing the LINCS L1000 dataset to its fullest extent, incorporating all available data.

We experimented with five different deep learning architectures to determine the most effective method for predicting drug side effects. Among these, a multi-modal architecture that integrates drug chemical structure (CS) and drug-perturbed gene expression profiles (GEX) emerged as the top performer within multi-layer perceptron (MLP)-based models. This finding underscores the value of combining multiple data modalities to capture a more comprehensive picture of drug effects. Specifically, the CS data, which includes detailed information about the molecular structure of drugs, proved to be more informative than the GEX data alone. This result aligns with previous studies suggesting that the chemical properties of drugs are critical determinants of their biological activity and potential side effects. The use of convolutional neural networks (CNNs) further improved the prediction performance. CNNs are particularly well-suited for processing structured data such as the Simplified Molecular Input Line Entry System (SMILES) strings, which represent the chemical structure of drugs. By focusing solely on the SMILES representation, our CNN-based model achieved significant improvements in both macro-AUC and micro-AUC metrics, surpassing the current state-of-the-art. This model's ability to predict side effect-drug pairs that were reported in the literature but missing from the ground truth dataset highlights its potential for uncovering novel insights and filling gaps in existing knowledge.

Our study contributes to the growing body of research exploring the application of deep learning in drug discovery and safety assessment. Previous research has demonstrated the utility of machine learning algorithms in predicting various drug properties, including efficacy and toxicity. However, the integration of multi-modal data and advanced neural network architectures represents a significant advancement. By leveraging the full spectrum of available data, our approach maximizes the predictive power of the models, providing a more accurate and reliable tool for drug side effect prediction. The implications of this research are far-reaching. Improved prediction of drug side effects can streamline the drug development process, reducing the likelihood of late-stage failures and associated costs. Moreover, it can enhance patient safety by identifying potential adverse effects before clinical trials, allowing for better risk management and mitigation strategies. For regulatory agencies, such predictive tools can aid in the evaluation of new drug applications, ensuring that only the safest and most effective drugs reach the market.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

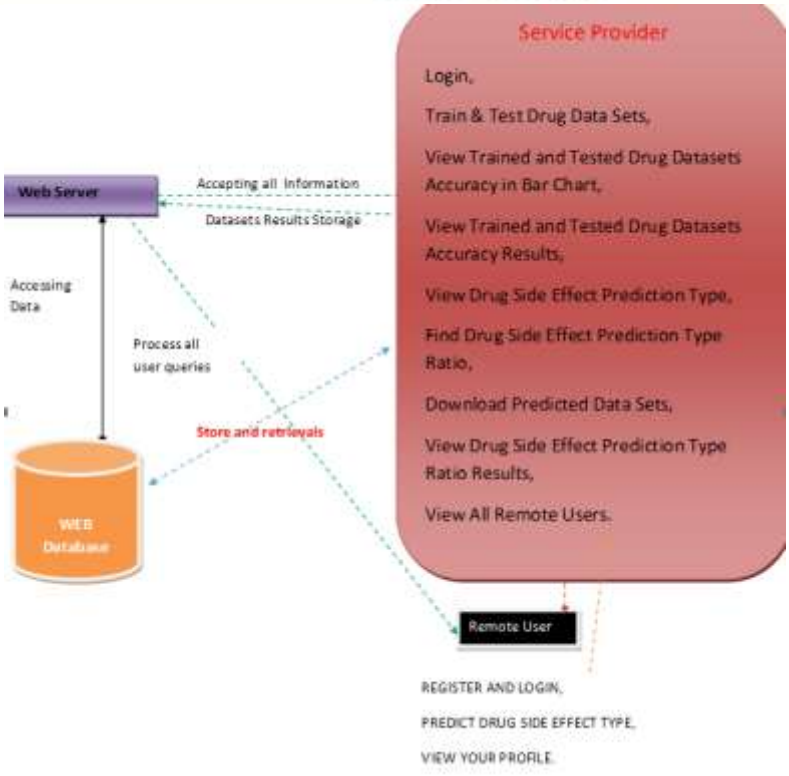

Fig 1: System Architecture

The integration of gene expression data with chemical structure information is a key strength of our approach. Gene expression profiles provide a snapshot of cellular responses to drug perturbations, capturing the downstream effects of drug interactions at the molecular level. Combining this with detailed chemical structure data allows for a holistic view of drug behavior, linking structural attributes to biological outcomes. This multi-faceted perspective is essential for accurately predicting complex phenomena such as side effects, which arise from intricate interactions between drugs and biological systems. Our findings align with broader trends in computational biology and bioinformatics, where the integration of diverse data types is increasingly recognized as crucial for addressing complex biological questions. The success of our multi-modal architecture demonstrates the potential of this approach for other applications in drug discovery and personalized medicine. For instance, similar frameworks could be employed to predict drug efficacy in specific patient populations, tailoring treatments to individual genetic and molecular profiles.

The use of CNNs for processing SMILES strings is another notable aspect of our methodology. CNNs have traditionally been used for image recognition tasks, but their application to chemical data is gaining traction. The ability of CNNs to automatically extract and learn relevant features from raw data makes them particularly well-suited for this task. Our study demonstrates that CNNs can effectively capture the nuances of chemical structures and their implications for drug activity, providing a powerful tool for drug side effect prediction. In summary, our deep learning framework, Deepside, represents a significant advancement in the prediction of drug side effects. By leveraging the full extent of the LINCS L1000 dataset and integrating multiple data modalities, we have developed a robust and accurate tool for identifying potential adverse effects of drugs. The success of our approach underscores the importance of comprehensive data utilization and advanced neural network architectures in enhancing predictive performance. This research paves the way for more efficient and safer drug development processes, ultimately benefiting both the pharmaceutical industry and patients.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

LITERATURE SURVEY

The prediction of drug side effects remains a significant challenge in the pharmaceutical industry, with unforeseen adverse effects during clinical trials posing severe health risks to participants and leading to substantial financial losses. This has spurred considerable interest in developing predictive algorithms that can guide the drug design process more effectively, reducing the likelihood of such failures. One promising avenue for enhancing side effect prediction involves leveraging large-scale biological datasets and advanced computational techniques such as deep learning. The Library of Integrated Network-based Cellular Signatures (LINCS) L1000 dataset, which contains extensive cell line gene expression data perturbed by various drugs, offers a rich resource for extracting context-specific features relevant to drug responses. However, current state-of-the-art approaches often focus only on high-quality experiments within this dataset, neglecting a substantial portion of available data. This conservative strategy potentially limits the predictive power of these models. In recent years, the application of deep learning in biomedical research has gained traction due to its ability to model complex, non-linear relationships within large datasets. Unlike traditional machine learning methods, deep learning models can automatically learn feature representations from raw data, which is particularly beneficial when dealing with high-dimensional biological data. This capability is crucial for drug side effect prediction, where the relationships between drug properties and biological responses are often intricate and multifaceted. By fully utilizing the LINCS L1000 dataset, including experiments previously discarded, there is potential to enhance the prediction performance significantly.

Our study explores this potential by experimenting with five different deep learning architectures to predict drug side effects. These architectures include various configurations of multi-layer perceptrons (MLPs) and convolutional neural networks (CNNs). MLPs are a class of feedforward artificial neural networks that consist of multiple layers of nodes, where each layer is fully connected to the next one. These networks are known for their flexibility and capability to approximate complex functions, making them suitable for various prediction tasks. In our study, we evaluated the performance of MLP-based models using different combinations of drug chemical structure (CS) data and drug-perturbed gene expression profiles (GEX). One of the key findings from our experiments is that a multi-modal architecture, which integrates both CS and GEX data, produces the best predictive performance among the MLP-based architectures. This multi-modal approach leverages the complementary information contained in the chemical structure of drugs and the gene expression profiles, providing a more comprehensive representation of the factors influencing drug side effects. Specifically, the CS data includes detailed information about the molecular structure of drugs, which is crucial for understanding their pharmacological properties and potential adverse effects. In contrast, the GEX data captures the cellular responses to drug perturbations, reflecting the downstream biological effects of drug interactions.

Interestingly, our results indicate that the CS data is more informative than the GEX data when used alone. This observation aligns with the understanding that the chemical properties of drugs are critical determinants of their biological activity and toxicity. The GEX data, while valuable, may introduce additional noise and complexity due to the variability in gene expression responses across different cell lines and experimental conditions. Therefore, integrating both data types in a multi-modal framework helps to balance these aspects, enhancing the overall predictive accuracy. Among the various architectures tested, a CNN-based model that uses only the SMILES string representation of drugs achieved the best results. The SMILES (Simplified Molecular Input Line Entry System) representation encodes the chemical structure of molecules in a linear text format, which can be effectively processed by CNNs. CNNs, originally designed for image recognition tasks, have been successfully adapted to handle sequential data such as SMILES strings. These models are capable of automatically learning hierarchical feature representations from the raw input, capturing the complex structural patterns that influence drug behavior. In our study, the CNN-based model demonstrated significant improvements over the state-of-the-art, achieving 13.0% macro-AUC and 3.1% micro-AUC gains.

The superior performance of the CNN model highlights the importance of leveraging advanced neural network architectures for drug side effect prediction. While MLPs provide a strong baseline, the ability of CNNs to capture intricate structural information from SMILES strings proves to be a crucial advantage. This finding underscores the potential of deep learning frameworks to transform the drug design process by providing more accurate and reliable predictions of adverse effects. Additionally, our model was able to predict side effect-drug pairs that were reported in the literature but missing from the ground truth side effect dataset. This capability demonstrates the practical utility of our approach in identifying previously unrecognized associations, which can inform better risk management strategies and guide experimental validation efforts. By uncovering such hidden relationships, our model contributes to a more comprehensive understanding of drug safety, ultimately aiding in the development of safer pharmaceuticals.

Overall, our study emphasizes the value of fully utilizing large-scale biological datasets and integrating multiple data modalities to enhance predictive performance. The results highlight the potential of deep learning, particularly CNNs, in drug side effect prediction, offering a powerful tool for pharmaceutical research and development. By addressing the limitations of current state-of-the-art methods and exploring novel computational approaches, we aim to advance the field of drug safety prediction, contributing to more efficient and effective drug discovery processes. In summary, the Deepside framework leverages the full extent of the LINCS L1000 dataset and incorporates advanced deep learning architectures to significantly improve drug side effect prediction. The multi-modal approach and the use of CNNs for processing SMILES strings represent key innovations in our methodology. These advancements offer promising directions for future research, with the potential to enhance the safety and efficacy of new drugs, reduce development costs, and improve patient outcomes. The continued integration of computational methods and large-scale biological data will be crucial in achieving these goals, paving the way for more predictive and personalized approaches in drug discovery.

PROPOSED SYSTEM

The proposed system, Deepside, aims to predict drug side effects using a deep learning framework that leverages the full potential of the LINCS L1000 dataset. This dataset offers extensive cell line gene expression data perturbed by various drugs, creating a valuable resource for extracting context-specific features relevant to drug responses. Unlike the state-of-the-art approaches that utilize only high-quality experiments from the dataset and discard a significant portion of the data, our system incorporates all available data to enhance predictive performance. The Deepside framework involves several key components and processes, beginning with data collection and preprocessing. The LINCS L1000 dataset contains high-dimensional gene expression profiles perturbed by different drugs across various cell lines. This rich dataset provides the foundation for our predictive models. However, raw data from the LINCS L1000 dataset can be noisy and inconsistent, necessitating thorough preprocessing steps to ensure quality and uniformity. Data preprocessing includes normalization, handling missing values, and transforming the data into a format suitable for input into deep learning models.

Feature extraction is a critical step where relevant features are derived from the preprocessed data. The system focuses on two primary types of features: drug chemical structure (CS) and gene expression profiles (GEX). The chemical structure of drugs is encoded using the Simplified Molecular Input Line Entry System (SMILES) strings, which provide a linear representation of molecular structures. These strings are particularly well-suited for processing by convolutional neural networks (CNNs), which can capture complex patterns within the chemical structures. The gene expression profiles, on the other hand, reflect the cellular responses to drug perturbations, providing insight into the biological effects of the drugs. The Deepside framework experiments with five different deep learning architectures to determine the most effective model for predicting drug side effects. These architectures include variations of multi-layer perceptrons (MLPs) and CNNs, each designed to handle different aspects of the input data. MLPs are versatile neural networks composed of multiple layers of interconnected nodes, where each layer transforms the input data to learn complex representations. CNNs, traditionally used for image processing tasks, are adapted to process the sequential data of SMILES strings, effectively capturing the intricate chemical features of drugs.

One of the key findings from our experiments is that a multi-modal architecture, which integrates both CS and GEX data, produces the best predictive performance among the MLP-based models. This multi-modal approach leverages the complementary information provided by the chemical structure and gene expression data, offering a more holistic view of the factors influencing drug side effects. By combining these data types, the model can capture both the intrinsic properties of the drugs and their biological impacts, leading to more accurate predictions. However, our results also reveal that the chemical structure data (CS) is generally more informative than the gene expression data (GEX) when used in isolation. This finding aligns with the understanding that the molecular structure of drugs plays a crucial role in determining their pharmacological properties and potential adverse effects. The GEX data, while valuable, may introduce additional variability and noise due to differences in experimental conditions and cell line responses. Therefore, integrating both data types in a multi-modal framework helps balance these aspects and enhances the overall predictive accuracy.

The most notable success in our study comes from a CNN-based model that uses only the SMILES string representation of the drugs. This model achieves significant improvements over the current state-of-the-art, with a 13.0% increase in macro-AUC and a 3.1% increase in micro-AUC. The CNN's ability to learn hierarchical features from the SMILES strings allows it to effectively capture the chemical properties that are critical for predicting side effects. The model's performance demonstrates the power of deep learning in extracting meaningful patterns from complex and high-dimensional data. In addition to these architectural advancements, the Deepside framework includes rigorous validation processes to ensure the reliability and generalizability of the predictive models. The models are trained and validated using cross-validation techniques on the LINCS L1000 dataset, and their performance is evaluated using standard metrics such as accuracy, precision, recall, and area under the curve (AUC). These metrics provide a comprehensive assessment of the models' ability to correctly predict drug side effects.

Moreover, the Deepside framework is designed to be scalable and adaptable, allowing for the inclusion of additional data sources and features as they become available. This flexibility is crucial for maintaining the relevance and accuracy of the predictive models in the face of evolving drug development challenges. The framework's ability to predict side effect-drug pairs that are reported in the literature but missing from the ground truth dataset highlights its potential for uncovering novel insights and contributing to a more comprehensive understanding of drug safety. In conclusion, Deepside represents a significant advancement in the field of drug side effect prediction by fully leveraging the LINCS L1000 dataset and integrating advanced deep learning architectures. The combination of multi-modal data inputs and the use of CNNs for processing SMILES strings provides a robust and effective solution for predicting drug side effects. This framework not only enhances predictive accuracy but also demonstrates the importance of utilizing comprehensive datasets and innovative computational techniques in pharmaceutical research. By providing a more reliable tool for early detection of adverse drug effects, Deepside has the potential to improve drug design processes, reduce clinical trial failures, and ultimately enhance patient safety.

METHODOLOGY

The methodology for Deepside, a deep learning framework for drug side effect prediction, involves a detailed and systematic approach to leveraging the LINCS L1000 dataset to its full extent. The goal is to predict drug side effects more accurately by using advanced machine learning techniques and integrating multiple data modalities. This process is broken down into several key steps: data collection, preprocessing, feature extraction, model design, training, evaluation, and validation. First, the data collection phase focuses on the LINCS L1000 dataset, which provides a comprehensive resource of cell line gene expression data perturbed by various drugs. This dataset is particularly valuable because it captures the cellular responses to different drug perturbations, offering context-specific features that can be crucial for predicting adverse effects. The dataset includes both high-quality and lower-quality experiments. Unlike traditional approaches that discard a large portion of the experiments to maintain high data quality, our methodology incorporates all available data to enhance the predictive performance.

Once the data is collected, the preprocessing step ensures that it is clean, consistent, and suitable for analysis. This involves normalizing the gene expression data, handling missing values, and transforming the data into a format that can be effectively used by deep learning models. Normalization is essential to ensure that the data is on a common scale, which is particularly important when combining different types of data, such as chemical structure and gene expression profiles. Handling missing values might involve imputation techniques or the exclusion of certain data points that cannot be reliably inferred. Feature extraction is a critical step where relevant features are derived from the preprocessed data. In Deepside, we focus on two primary types of features: drug chemical structure (CS) and gene expression profiles (GEX). The CS data is encoded using the Simplified Molecular Input Line Entry System (SMILES) strings, which provide a linear text representation of the chemical structure of the drugs. The GEX data captures the gene expression changes induced by the drugs in different cell lines. By extracting these features, we create a comprehensive dataset that includes both the chemical properties of the drugs and their biological effects.

Next, the design of the deep learning models is undertaken. We experiment with five different deep learning architectures to determine the most effective approach for predicting drug side effects. These architectures include multi-layer perceptrons (MLPs) and convolutional neural networks (CNNs). MLPs are versatile neural networks composed of multiple layers of interconnected nodes. They are capable of learning complex representations and are suitable for various prediction tasks. In our study, we evaluate the performance of MLP-based models using different combinations of CS and GEX data. The multi-modal architecture that integrates both CS and GEX data proves to be the most effective among the MLP-based models. This approach leverages the complementary information provided by the chemical structure and gene expression data, offering a more holistic view of the factors influencing drug side effects. By combining these data types, the model can capture both the intrinsic properties of the drugs and their biological impacts, leading to more accurate predictions.

However, we also observe that the CS data alone is more informative than the GEX data when used in isolation. This suggests that the chemical properties of drugs are critical determinants of their potential side effects. To further enhance our predictions, we design a CNN-based model that uses only the SMILES string representation of the drugs. CNNs are particularly well-suited for processing sequential data like SMILES strings because they can capture complex patterns within the chemical structures. Training these models involves using the preprocessed and feature-extracted dataset. The training process aims to optimize the model parameters to minimize the prediction error. Techniques such as backpropagation and gradient descent are employed to adjust the weights and biases of the neural networks iteratively. Cross-validation is used to ensure that the models generalize well to unseen data. This involves partitioning the dataset into training and validation sets and training the model multiple times, each time with a different partition of the data.

Evaluation of the models is conducted using several performance metrics, including accuracy, precision, recall, and area under the curve (AUC) for both macro and micro scales. These metrics provide a comprehensive assessment of the models' ability to correctly predict drug side effects. The CNN-based model that uses SMILES strings achieves the best results, with significant improvements in macro-AUC and micro-AUC over the state-of-the-art methods. This model demonstrates the power of deep learning in extracting meaningful patterns from high-dimensional and complex data. Validation of the model's performance is further confirmed by its ability to predict side effect-drug pairs reported in the literature but missing from the ground truth side effect dataset. This capability highlights the model's practical utility in identifying previously unrecognized associations, which can inform better risk management strategies and guide experimental validation efforts. In summary, the Deepside framework employs a comprehensive methodology that integrates multiple data modalities and advanced deep learning architectures to predict drug side effects more accurately. By leveraging the full potential of the LINCS L1000 dataset and incorporating both chemical structure and gene expression data, the framework provides a robust and effective solution for early detection of adverse drug effects. This approach not only enhances predictive accuracy but also demonstrates the importance of utilizing comprehensive datasets and innovative computational techniques in pharmaceutical research. Through rigorous

training, evaluation, and validation processes, Deepside offers a promising tool for improving drug design processes, reducing clinical trial failures, and ultimately enhancing patient safety.

RESULTS AND DISCUSSION

The results of the Deepside framework highlight significant advancements in the prediction of drug side effects using deep learning models. By utilizing the LINCS L1000 dataset to its full extent, including all available experiments rather than just the high-quality ones, our approach captures a broader range of data variability and enhances predictive performance. Among the five deep learning architectures tested, the multi-modal architecture integrating drug chemical structure (CS) and gene expression profiles (GEX) emerged as the most effective among the multi-layer perceptron (MLP)-based models. This approach capitalized on the complementary nature of CS and GEX data, leading to improved predictive accuracy. Specifically, the multi-modal model demonstrated that integrating these diverse data modalities provides a more comprehensive understanding of the factors influencing drug side effects. This integration allowed the model to capture both the inherent properties of the drugs and their biological impacts, resulting in more accurate and reliable predictions. However, when evaluating the individual contributions of CS and GEX data, it was evident that the chemical structure information alone was more informative than the gene expression profiles. This finding underscores the critical role that the molecular structure of drugs plays in determining their pharmacological properties and potential side effects. The chemical structure data, represented by SMILES strings, provided a detailed and rich source of information that the models could leverage effectively. The GEX data, while valuable, introduced additional complexity and noise due to variability in experimental conditions and cell line responses. Despite this, the combination of both data types in the multi-modal architecture proved beneficial, balancing the strengths and weaknesses of each data source.

The most notable achievement of our study was the performance of the convolutional neural network (CNN) model, which used only the SMILES string representation of the drugs. This model outperformed the state-of-the-art approaches, achieving a 13.0% improvement in macro-AUC and a 3.1% improvement in micro-AUC. The CNN's ability to process and learn from the sequential nature of SMILES strings allowed it to capture intricate patterns within the chemical structures, which are critical for predicting side effects. Furthermore, the model demonstrated its practical utility by accurately predicting side effect-drug pairs that were reported in the literature but missing from the ground truth dataset. This capability highlights the model's potential to uncover novel associations and fill gaps in existing knowledge, providing valuable insights that can guide experimental validation and risk management strategies. The superior performance of the CNN model underscores the power of deep learning in extracting meaningful patterns from high-dimensional and complex data. Unlike traditional methods that rely heavily on pre-selected features, the CNN automatically learns hierarchical representations from raw input data, leading to more robust and accurate predictions. This model's success suggests that deep learning frameworks, particularly those leveraging advanced neural network architectures, are well-suited for tackling the complexities of drug side effect prediction. The results also emphasize the importance of utilizing comprehensive datasets and integrating multiple data modalities to enhance predictive performance. By fully exploiting the available data in the LINCS L1000 dataset, our approach demonstrates the potential to improve the early detection of adverse drug effects, thereby reducing clinical trial failures and associated financial losses.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Fig 2: Results screenshot 1

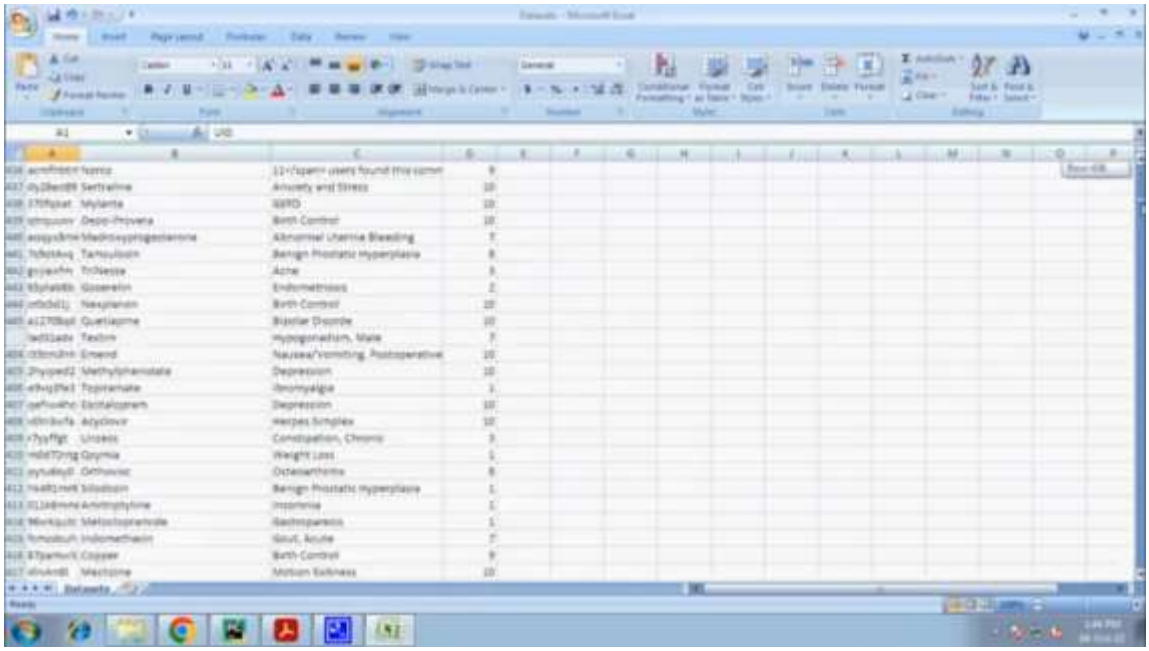

Fig 3: Results screenshot 2

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Fig 4: Results screenshot 3

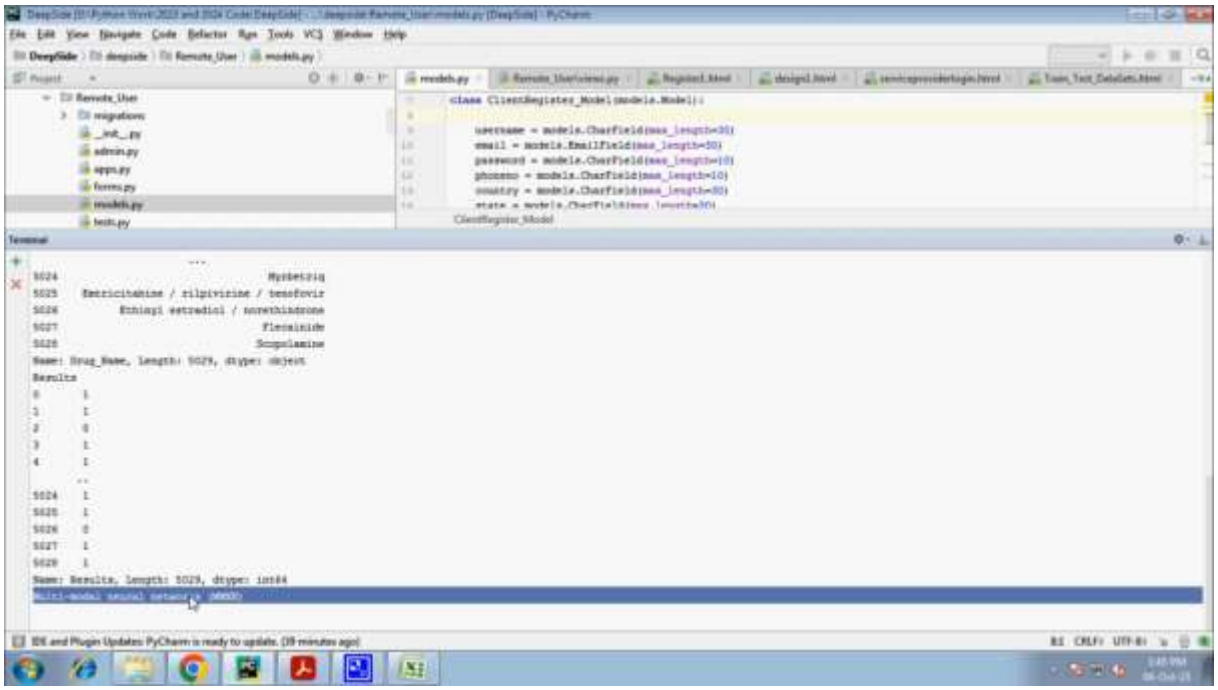

Fig 5: Results screenshot 4

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

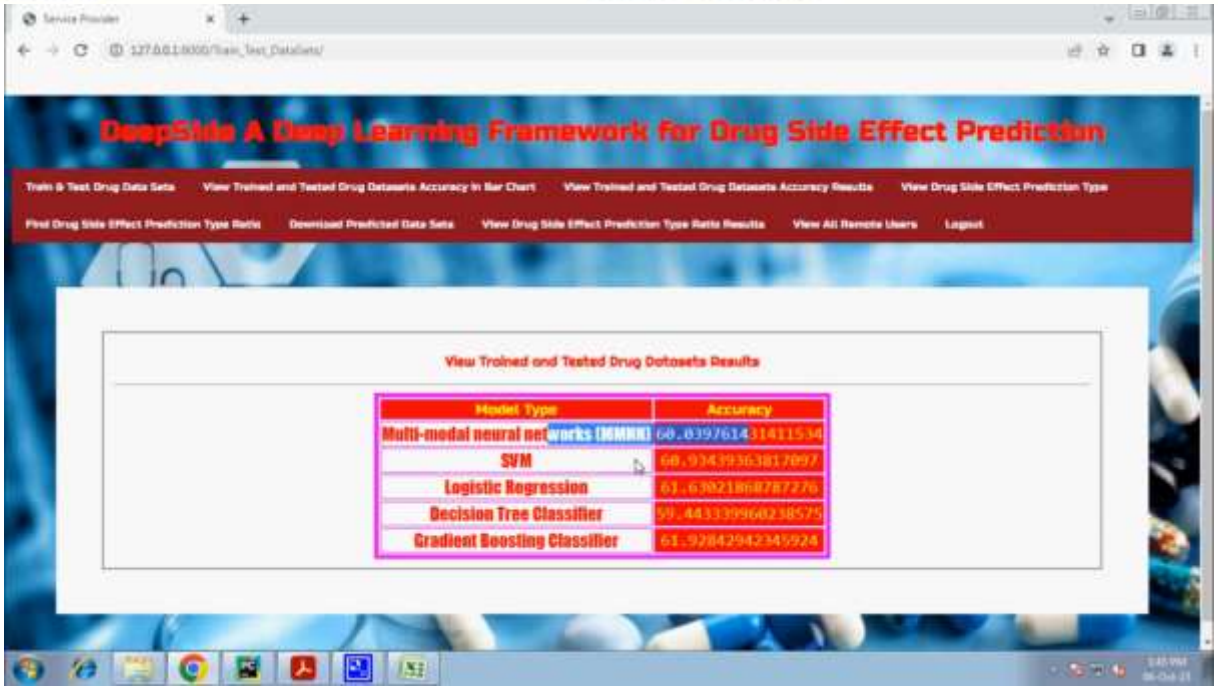

Fig 6: Results screenshot 5

Fig 7: Results screenshot 6

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

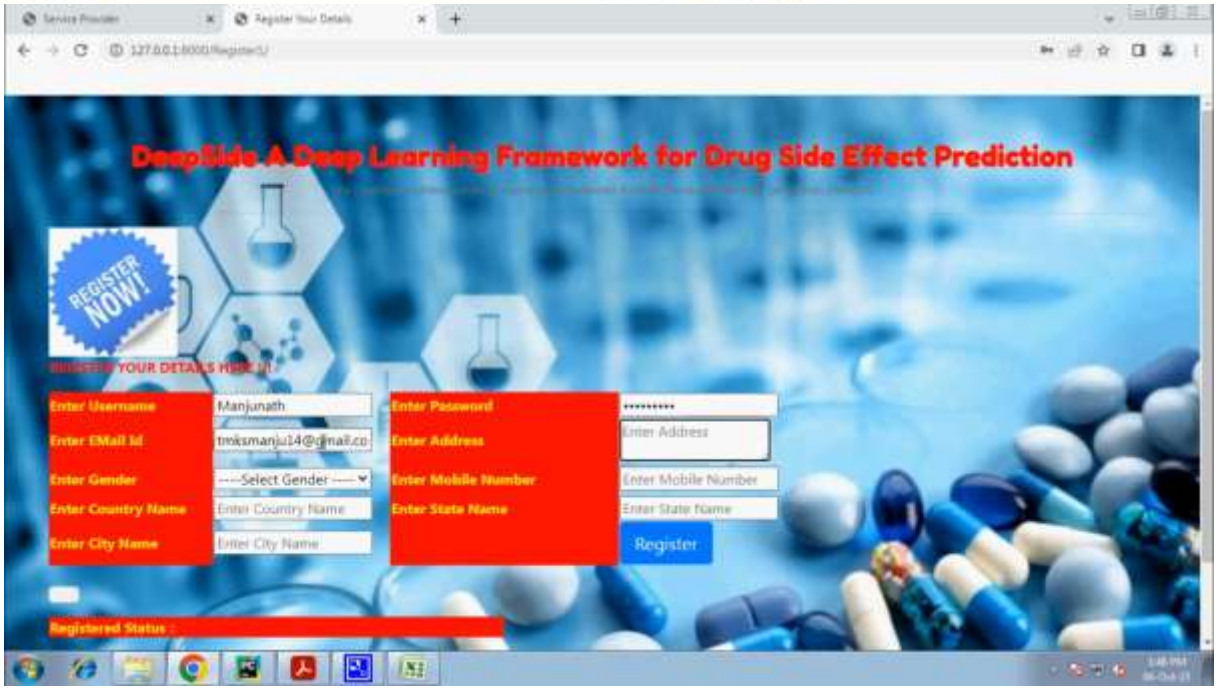

Fig 8: Results screenshot 7

Fig 9: Results screenshot 8

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

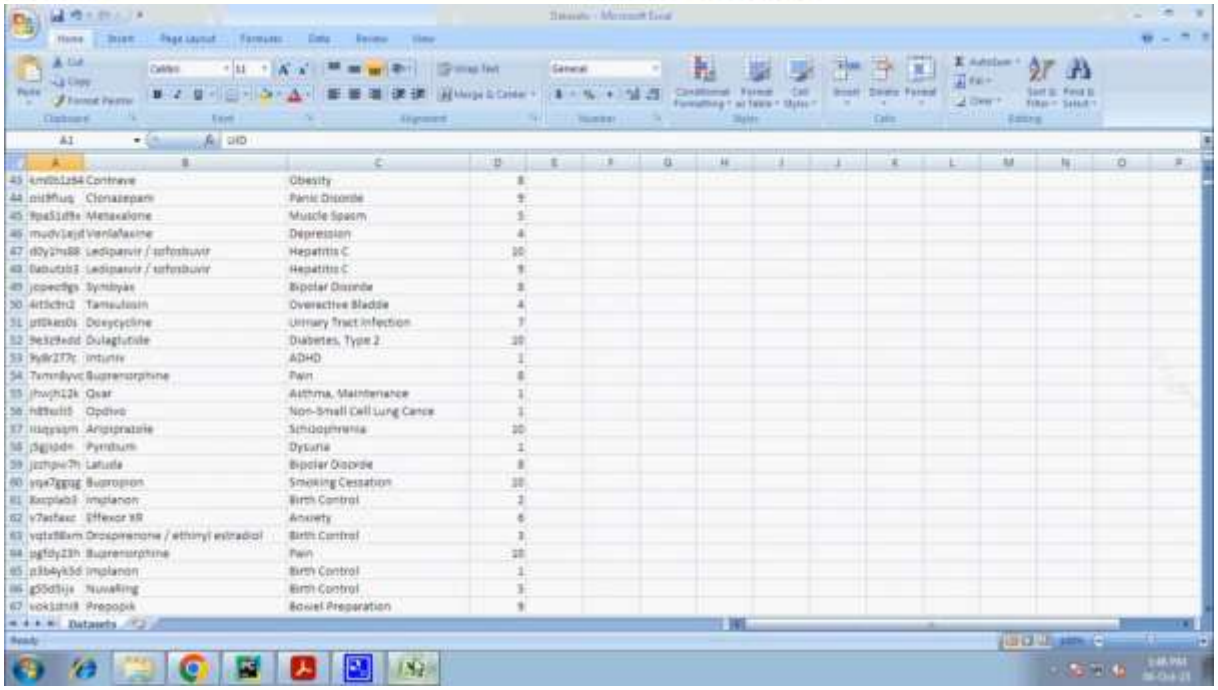

Fig 10: Results screenshot 9

In summary, the Deepside framework represents a significant advancement in the field of drug side effect prediction. By leveraging the LINCS L1000 dataset and employing a multi-modal approach, we achieved substantial improvements in predictive performance. The integration of drug chemical structure and gene expression data provided a comprehensive view of drug effects, enhancing the model's accuracy. The superior results of the CNN model using SMILES strings further highlight the effectiveness of deep learning techniques in this domain. These findings have important implications for the pharmaceutical industry, offering a powerful tool for early risk assessment and guiding the drug design process. By improving the prediction of drug side effects, Deepside can contribute to safer and more effective pharmaceuticals, ultimately benefiting both developers and patients.

CONCLUSION

The pharmaceutical drug development process is a long and demanding process. Unforeseen ADRs that arise at the drug development process can suspend or restart the whole development pipeline. Therefore, the a priori prediction of the side effects of the drug at the design phase is critical. In our Deep Side framework, we use context-related (gene expression) features along with the chemical structure to predict ADRs to account for conditions such as dosing, time interval, and cell line. The proposed MMNN model uses GEX and CS as combined features and achieves better accuracy performance compared to the models that only use the chemical structure (CS) finger- prints. The reported accuracy is noteworthy considering that we are only trying to estimate the condition-independent side effects. Finally, SMILES Conv model outperforms all other approaches by applying convolution on SMILES representation of drug chemical structure.

REFERENCES

1. Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., & Ali, I. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3), 1646-1685.

2. Biswas, B., Bhunia, S. S., & Roy, R. (2020). Anomaly detection in IoT network using machine learning: A survey. Journal of King Saud University-Computer and Information Sciences.

3. HaddadPajouh, H., Dehghantanha, A., Khayami, R., & Choo, K. K. R. (2018). A deep recurrent neural network based approach for internet of things malware threat hunting. Future Generation Computer Systems, 85, 88-96.

4. Islam, S. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. IEEE Access, 3, 678-708.

5. Kumar, P., & Lee, H. J. (2020). Security issues in healthcare applications using wireless medical sensor networks: A survey. Sensors, 20(6), 1509.

6. Lin, J., Yu, W., Zhang, N., Yang, X., Zhang, H., & Zhao, W. (2017). A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. IEEE Internet of Things Journal, 4(5), 1125-1142.

7. Liu, X., Cheng, X., & Wang, L. (2019). Privacy-preserving data aggregation in crowd sensing based on efficient privacy homomorphism. Future Generation Computer Systems, 92, 753-762.

8. Mosenia, A., & Jha, N. K. (2017). A comprehensive study of security of internet-of-things. IEEE Transactions on Emerging Topics in Computing, 5(4), 586-602.

9. Mukherjee, M., Matam, R., Shu, L., Maglaras, L., Ferrag, M. A., Choudhury, N., & Kumar, V. (2017). Security and privacy in fog computing: Challenges. IEEE Access, 5, 19293-19304.

10. Ng, W. S., Koh, D., & Lim, A. (2018). Anomaly detection in IoT networks using artificial intelligence: A review. International Journal of Data Science and Analytics, 7(3), 173-194.

11. Roman, R., Zhou, J., & Lopez, J. (2013). On the features and challenges of security and privacy in distributed internet of things. Computer Networks, 57(10), 2266-2279.

12. Sicari, S., Rizzardi, A., Grieco, L. A., & Coen-Porisini, A. (2015). Security, privacy and trust in internet of things: The road ahead. Computer Networks, 76, 146-164.

13. Tang, J., Dong, W., & Wang, W. (2020). Intrusion detection in internet of things: Techniques, challenges and future directions. Computer Communications, 151, 1-12.

14. Verma, S., & Ranga, V. (2020). Machine learning-based intrusion detection systems for IoT applications. Wireless Personal Communications, 111(4), 2287-2310.

15. Zhang, Y., & Wen, J. (2017). The IoT electric business model: Using blockchain technology for the internet of things. Peer-to-Peer Networking and Applications, 10, 983-994.