International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

AI/ML Based Disease Prediction

Prof. Yogesh Handge
Dept. of Computer Engineering Pune
Institute of Computer Technology
Pune, India
yahandge@pict.edu

Sarthak Dhaytonde
Dept. of Computer Engineering
Pune Institute of Computer Technology
Pune, India
sarthakdhaytonde014@gmail.com

Ajit Kale
Dept. of Computer Engineering Pune
Institute of Computer Technology
Pune, India
ajitkale2406@gmail.com

Soham Labba
Dept. of Computer Engineering Pune Institute of Computer Technology
Pune, India
labbasoham18@gmail.com

Kartik Kasrewar
Dept. of Computer Engineering Pune Institute of Computer Technology
Pune, India kasrewarkartik.0709@gmail.com

***Abstract*—Symptom- grounded complaint vaticination has come in- creasingly important in healthcare due to the growing need for accurate, early judgments and substantiated treatment plans. Traditional opinion styles calculate heavily on mortal moxie, which can be time- consuming and prone to crimes, especially with complex or coinciding symptoms. The integration of machine literacy(ML) and deep literacy(DL) ways has significantly enhanced automated complaint vaticination systems, enabling briskly and more accurate prognostications. still, chal- lenges remain, particularly in handling deficient symptom data, perfecting vaticination delicacy for rare conditions, and icing model interpretability. This check paper reviews the current state of exploration in symptom- grounded complaint vaticination models, including machine literacy algorithms, cold-blooded approaches, and data preprocessing ways. It also explores recent advance- ments in using neural networks and natural language processing NLP) for symptom analysis. also, the check proposes a frame that combines multi-modal data emulsion, bettered algorithmic performance, and resolvable AI to give further de- pendable and interpretable prognostications. The check examines the technological inventions, challenges, and unborn operations of this approach in clinical decision support systems.**

I. INTRODUCTION

In healthcare, early and accurate complaint opinion is cru- cial for effective treatment and bettered patient issues. Tradi- tional individual styles frequently calculate on mortal moxie, which can be time- consuming, prone to crimes, and incon- sistent across different interpreters. As the complexity of med- ical conditions increases, homemade opinion becomes indeed more grueling , especially when multiple symptoms lap across different conditions. To address these limitations, bus- slept complaint vaticination systems grounded on machine literacy ML) and artificial intelligence(AI) have gained elevation. These systems influence data- driven approaches to dissect patient symptoms and induce complaint prognostications, but several challenges remain undetermined

1) Deficient and Noisy Symptom Data Medical datasets frequently contain missing or inconsistent symptom informa- tion, which can negatively affect the delicacy of vaticination models.

2) Handling Rare conditions numerous vaticination systems strug- gle to directly prognosticate rare conditions due to the lack of suf- ficient data and the imbrication of symptoms with further common conditions.

3) Interpretability of prognostications In healthcare, it's pivotal for automated systems to give not only accurate prog- nostications but also interpretable and transparent logic behind the prognostications to gain the trust of medical professionals.

II. LITERATURE SURVEY

A. Machine Learning for Symptom-based Disease Prediction

Machine Learning for Symptom- grounded Disease Predic- tion Machine literacy(ML) has been extensively applied to symptom- grounded complaint vaticination, aiming to amelio- rate diag- nostic delicacy and speed. Traditional ML styles, similar as decision trees and arbitrary timbers, have been used to classify conditions grounded on symptom datasets. For case, Singh and Kumar reviewed colorful ML ways for complaint opinion, pressing the effectiveness of algorithms similar as support vector machines(SVMs) and k- nearest neighbors(KNN). still, they noted that these styles frequently struggle with unstable data and complex,multi-symptom conditions. The preface of deep literacy(DL) ways has allowed more so- phisticated models to be developed for complaint vaticination. Deep neural networks(DNNs) and convolutional neural net- works(CNNs) have demonstrated significant im- provements in assayingmulti-dimensional symptom data. Sharma bandied the use of CNNs in relating patterns in symptom clusters that were preliminarily delicate to capture with traditional styles. Despite these advances, DL models frequently bear large datasets and are prone to overfitting when applied to small or imbalanced datasets.

B. Handling Incomplete and Noisy Symptom Data

Medical datasets are constantly deficient, containing missing values or noisy data, which can significantly affect the per- formance of complaint vaticination models. Patel and Mehta

proposed insinuation ways combined with ML models to handle missing data in clinical settings. They showed that combining k-means clustering with insinuation bettered the delicacy of vaticination systems.

A mongrel approach using insinuation and anomaly discov- ery was presented by Goyal Their system employed autoen- coders to reconstruct missing symptom data and linked outliers that could distort the vaticination delicacy. Although this ap- proach enhances data quality, it still faces challenges when dealing with large- scale datasets from multiple sources.

C. Rare Disease Prediction

Predicting rare conditions poses a significant challenge due to the lack of sufficient training data. Traditional ML models per- form inadequately in these cases, frequently misclassify- ing rare conditions due to their underrepresentation in datasets. Singh proposed an ensemble literacy approach that balances the dataset by oversampling rare complaint cases. This system, while effective in perfecting the discovery of rare conditions, still struggles with generalizing to unseen cases.

More lately, transfer literacy ways have been intro- duced to address the failure of data in rare complaint vaticination. Gupta employed transfer literacy by usingpre-trained models on large medical datasets to ameliorate rare complaint bracket. Although promising, the effectiveness of transfer literacy is limited by the vacuity of large, well- annotated medical datasets.

D. Explainability and Interpretability in Disease Prediction Models

Explainability is a critical demand in healthcare, where professionals need to trust and understand the prognostications made by automated systems. Black- box models, similar as deep neural networks, are frequently delicate to interpret, making them less desirable in clinical settings. Saha proposed the use of decision trees combined with neural networks to produce an interpretable vaticination frame. This mon- grel model offered bettered interpretability, but at the cost of reduced vaticination delicacy compared to more complex models. resolvable AI(XAI) ways, similar as LIME(Original Interpretable Model- agnostic Explanations) and SHAP(Shap- ley cumulative Explanations), have gained fashionability in making complex models more interpretable. Mishra applied SHAP to a symptom- grounded vaticination model, enabling clinicians to visu- alize the donation of each symptom to the final vaticination. This approach enhanced translucency but was computationally expensive for real- time use in clinical surroundings.

1) Multi-Modal Data Fusion in Disease Prediction: Mod- ern complaint vaticination systems decreasingly incorporate data from multiple sources, similar as patient history, inherita- ble infor- mation, and imaging data, alongside symptom data. Integrating these different data types presents a significant challenge. Das introduced amulti-modal literacy frame that combines textbook- grounded symptom data with imaging data to ameliorate complaint prognostications. They used a emulsion of CNNs for image analysis and intermittent neural networks(RNNs) for textual data. This system showed ad- vancements in vaticination delicacy, particu- larly for complex conditions with different symptoms, but needed expansive computational coffers. A analogous approach was employed by Wang., who proposed a frame that integrates electronic health records(EHR) with symptom data to enhance complaint vaticination models. Their results demonstrated the benefits ofmulti-modal data emulsion in perfecting individual perfec- tion, though the system plodded with the real- time processing of high- dimensional data.

III. METHODOLOGIES STUDIED

A. Traditional Machine Learning Techniques

Methods like decision trees, support vector machines (SVM), and k-nearest neighbors (KNN) were used to classify diseases based on symptom data.

1) Limitations: These methods struggled with unbalanced datasets and complex diseases involving overlapping symp- toms.

B. Data Imputation for Incomplete Symptom Data

Techniques like k-means clustering and autoencoders were used to handle missing or noisy data through imputation and anomaly detection.

1) Limitations: These methods improve data quality but face challenges with large-scale, heterogeneous datasets.

IV. PROPOSED METHODOLOGY

The proposed complaint vaticination system grounded on symptoms integrates colorful advanced machine literacy ways, data preprocessing, andmulti-modal literacy to effectively ad- dress the challenges linked in the literature:

A. Symptom Data Preprocessing and Imputation

The system preprocesses symptom data by handling missing or deficient inputs using data insinuation ways similar as k- means clustering and autoencoders. These styles fill gaps in the data to insure a more complete dataset for analysis. also, noise reduction ways are employed to clean the data, making it more dependable for vaticination.

B. Machine Learning-Based Disease Prediction

The core element of the system uses an ensemble of ma- chine literacy models, similar as Random timbers, SVM, and CNNs, to prognosticate conditions grounded on symptoms. By using ensemble literacy, the system improves bracket perfor- mance and addresses issues of rare complaint vaticination. Addi- tionally, deep literacy models are employed to capture com- plex connections between symptoms that traditional styles might overlook.

C. *Real-Time Prediction and Feedback*

The system is designed to reuse symptom inputs in real- time, furnishing immediate prognostications and implicit com- plaint judgments . The real- time feedback circle enables timely medical interventions and updates the model as further patient data be- comes available, allowing the system to con- tinuously ameliorate its prognostications through underpinning literacy.

V. CONCLUSION AND FUTURE SCOPE

This check paper reviews the current state of exploration in au- tomated symptom- grounded complaint vaticination and highlights crucial advancements in machine literacy algo- rithms, data prepro- cessing ways, and model interpretability. The proposed system addresses critical challenges in handling deficient symptom data, perfecting the vaticination of rare conditions, and icing the interpretability of prognostications. By leverag- ing advanced data- driven approaches, the system enhances the delicacy and trustability of automated complaint vaticination, furnishing precious support to healthcare profes- sionals. While challenges remain, similar as integratingmulti- modal data and perfecting real- time vaticination performance, the system offers a promising result to the complications of automated complaint opinion.

Looking ahead, unborn work could concentrate on expand- ing the system's capabilities to handle larger, more different datasets, including case histories, inheritable information, and real- time health monitoring data. also, perfecting the model's capability to acclimatize to new conditions and integrating further sophis- ticated resolvable AI ways could further in- crease its utility in clinical settings. The integration of pall- grounded and allied literacy approaches also presents openings to enhance scalability and insure patient data sequestration while maintaining model delicacy across different healthcare institutions.

REFERENCES

[1] G. Singh, R. Kumar, *A Survey of Machine Learning Techniques for Dis- ease Diagnosis*, In Proceedings of the IEEE 6th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE Press, 868–874.

[2] Y. Zhang, J. Liu, F. Xiao, *Symptom-based Diagnosis of Complex Dis- eases Using Machine Learning Techniques*, In Proceedings of the IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS '19). IEEE Press, 883–968.

[3] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, *Disease Prediction Using Machine Learning Over Big Data from Healthcare Communities*, IEEE Access, 5, 8869–8879.

[4] R. B. S, S. D. S, A. Shabna, *Early Prediction of Diseases Using Machine Learning Techniques*, In Proceedings of the IEEE International Confer- ence on Computational Intelligence and Computing Research (ICCIC '20). IEEE Press, 904–3104.

[5] M. Tripathi, A. Rathore, P. Upadhyay, *Symptom-Based Disease Prediction Using Decision Tree and Random Forest Algorithms*, International Journal of Computer Applications (IJCA), Vol. 175, No. 6, 26–29.

[6] Z. Z. Khan, M. Imran, M. Ghamdi, *A Comprehensive Survey on Machine Learning for Healthcare*, IEEE Access, 8, 50150–50172.

[7] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, *Deep Learning for Health- care: Review, Opportunities, and Challenges*, Journal of Medical Internet Research (JMIR), 19, No. 5, e144.

[8] A. Rajan, K. Ramesh, *Machine Learning Approaches for Symptom-Based Disease Prediction: A Review*, Procedia Computer Science, Vol. 165, 94–103.

[9] B. Jain, A. Sharma, A. Mittal, *Prediction of Disease Using Machine Learning Algorithms in Healthcare*, International Journal of Emerging Technologies in Engineering Research (IJETER), Vol. 8, No. 1, 32–35.

[10] S. M. S, J. Prabha, *Medical Disease Prediction Using Support Vector Machine and K-Nearest Neighbor*, Journal of Innovations in Computer Science and Engineering (JIISCE), Vol. 9, No. 2, 82–88.