e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

PATIENT CASE SIMILARITY

Mahima[*1], VB Jeevinee[*2], Mohammed Salman Khan[*3]
[*1,2,3]Presidency University, School Of Computer Science And Engineering, Bangalore, India.
DOI : https://www.doi.org/10.56726/IRJMETS48246

ABSTRACT

The Patient Case Similarity project revolutionizes healthcare analytics, employing advanced data science for personalized and data-driven interventions. Utilizing sophisticated algorithms, it identifies patient case similarities, enhancing diagnostic accuracy, tailoring treatments, and improving outcomes. Beyond individualized care, it offers a holistic view of disease trajectories, aiding complication anticipation, optimizing resource allocation, and advancing medical research. This project has the potential to transform clinical decision-making, empowering clinicians with a robust tool for navigating complex patient data. In essence, it redefines healthcare delivery by unlocking latent knowledge, showcasing the transformative power of data science in shaping patient-centered care.

I. INTRODUCTION

In the dynamic healthcare landscape, marked by rising patient numbers and escalating emergencies, the demand for robust solutions is paramount. Our proposed model, intricately designed for navigating complex challenges without immediate medical treatment, offers a beacon of hope with a prompt, individualized approach to care. At its core, the model identifies patients similar to a specified index patient by meticulously exploring past medical records. This groundbreaking method provides tailored forecasts, significantly elevating patient care quality, crucial in dire circumstances. Surpassing traditional methods, it meticulously examines health data, forecasting patients' future health statuses using advanced computer-based techniques. In emergencies, where quick medical treatment is limited, the model becomes indispensable, optimizing reactions by empowering healthcare practitioners with well-informed decisions based on comparable patient records. This model heralds a new era in patient-centric care, laying the foundation for transformative progress in predictive healthcare analytics, revolutionizing personalized, data-driven medical interventions.

II. METHODOLOGY

The proposed Flask web app integrates modern web development, machine learning (ML), and natural language processing (NLP) to address healthcare tech gaps, enhancing medication recommendations and disease predictions. ML algorithms, like Random Forests or Neural Networks, decode intricate health data trained on extensive patient datasets. NLP analyzes patient-reported symptoms, transforming natural language inputs for precise ML interpretation. The Flask UI ensures accessibility for users with varying technical expertise, seamlessly connecting with ML and NLP for real-time predictions in healthcare.

Data Acquisition and Preprocessing

Collect precise patient data from reputable medical sources. Clean and normalize the dataset meticulously, filling blanks, removing duplication, and standardizing formats. This enhances accuracy, facilitating improved ML and NLP analysis.

Machine Learning Model Development

Start by carefully selecting task-appropriate ML algorithms, such as SVM, Random Forests, and Neural Networks, based on dataset characteristics. Proceed with model training using preprocessed data, dividing it into testing and training sets. Evaluate models for accuracy, precision, and recall, fine-tuning them for improved performance based on evaluation findings.

Natural Language Processing for Symptom Analysis

In symptom text processing, raw user- inputted symptoms undergo normalization and NLP techniques like tokenization. Feature extraction, utilizing methods like TF-IDF, follows, converting text data into numerical values for ML analysis. Integrated with ML models, this approach ensures accurate disease predictions from NLP-processed data.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

Development of the Flask Web Application

Flask is chosen for its ease in rapidly building web apps, seamlessly deploying ML and NLP models. Prioritizing user-friendly design, it focuses on accessible forms and device responsiveness.

Validation and Testing

The project evaluates ML models for prescription recommendation and illness prediction, ensuring real-world capability through accuracy assessments. User input and robustness testing validate correctness, responsiveness, and stability. Cross-browser and cross-platform testing guarantee reliability.

III. MODELING AND ANALYSIS

Utilizing advanced ML, this healthcare analytics system predicts patient similarity, addressing challenges through preprocessing, neural networks, and Flask integration.

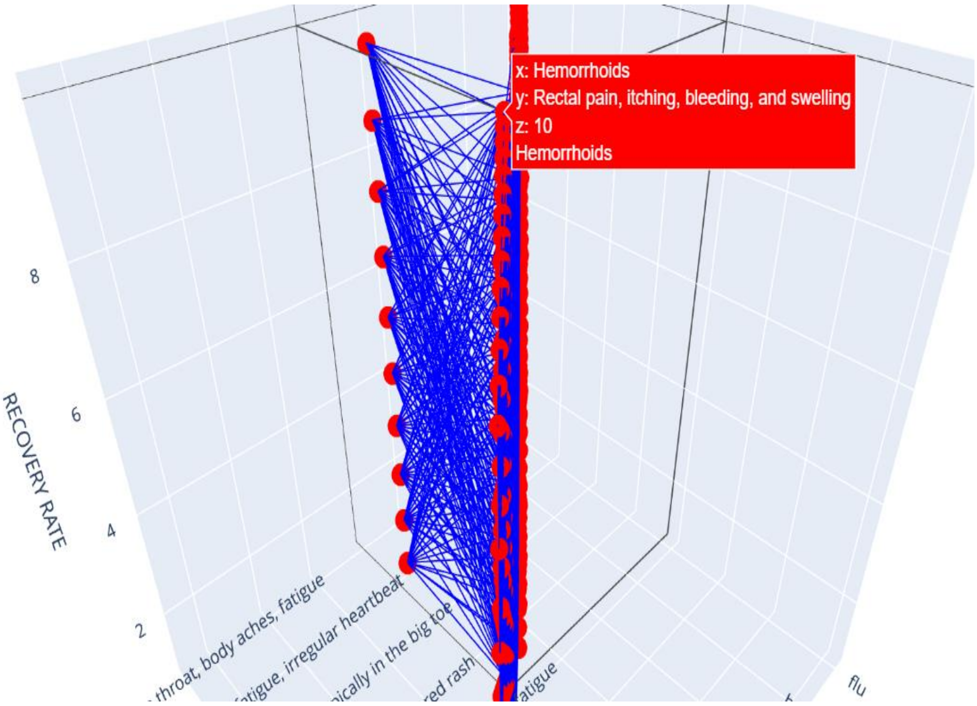

Figure 1: 3D Scatter Plot

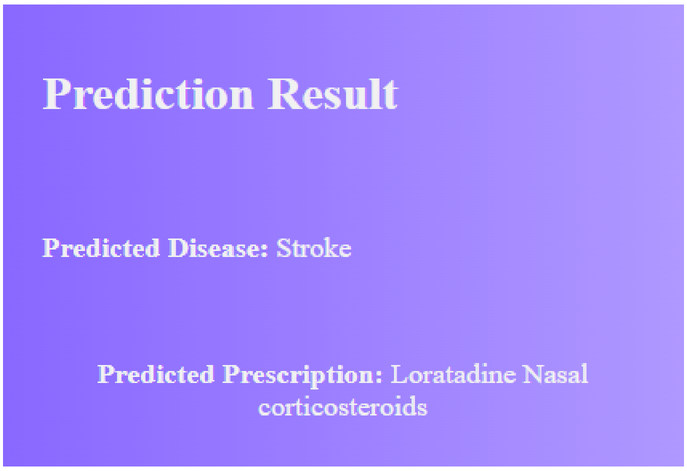

Figure 2: Prediction

IV. RESULTS AND DISCUSSION

In the healthcare realm, our predictive patient case similarity system revolutionizes diagnostics and patient care. Utilizing advanced machine learning with a user-friendly interface, it refines input data and predicts diseases accurately. Ethical data gathering prioritizes privacy, and an iterative loop ensures continuous improvement. The system seamlessly integrates into clinical settings, supporting scalability and accessibility. It

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:01/Jauuary-2024 Impact Factor- 7.868 www.irjmets.com

addresses challenges in predictive healthcare analytics, particularly in emergency scenarios. Its multifaceted approach, including privacy-preserving techniques, neural network models, and Flask web application integration, showcases a commitment to ethical considerations and predictive accuracy. Overall, it heralds a new era of patient-centric care, leveraging cutting-edge technologies to meet dynamic healthcare demands.

Table 1. Accuracy of both the models

SN.	Prediction Model	Accuracy
1	Disease	0.9578
2	Prescription	0.6127

```
Epoch 1/10
2439/2439 [==============================] - 7s 2ms/step - loss: 0.2077 - accuracy: 0.9578 - val_loss: 0.0474 - val_accuracy: 0.9672
Epoch 2/10
2439/2439 [==============================] - 5s 2ms/step - loss: 0.0479 - accuracy: 0.9664 - val_loss: 0.0475 - val_accuracy: 0.9646
Epoch 3/10
2439/2439 [==============================] - 5s 2ms/step - loss: 0.0471 - accuracy: 0.9659 - val_loss: 0.0473 - val_accuracy: 0.9672
Epoch 4/10
2439/2439 [==============================] - 6s 3ms/step - loss: 0.0469 - accuracy: 0.9656 - val_loss: 0.0476 - val_accuracy: 0.9672
Epoch 5/10
2439/2439 [==============================] - 5s 2ms/step - loss: 0.0467 - accuracy: 0.9661 - val_loss: 0.0480 - val_accuracy: 0.9646
Epoch 6/10
2439/2439 [==============================] - 6s 3ms/step - loss: 0.0465 - accuracy: 0.9671 - val_loss: 0.0474 - val_accuracy: 0.9646
Epoch 7/10
2439/2439 [==============================] - 5s 2ms/step - loss: 0.0466 - accuracy: 0.9665 - val_loss: 0.0474 - val_accuracy: 0.9672
Epoch 8/10
2439/2439 [==============================] - 7s 3ms/step - loss: 0.0466 - accuracy: 0.9661 - val_loss: 0.0473 - val_accuracy: 0.9672
Epoch 9/10
2439/2439 [==============================] - 5s 2ms/step - loss: 0.0464 - accuracy: 0.9671 - val_loss: 0.0475 - val_accuracy: 0.9646
Epoch 10/10
2439/2439 [==============================] - 6s 2ms/step - loss: 0.0464 - accuracy: 0.9666 - val_loss: 0.0473 - val_accuracy: 0.9646
```

Figure 3: Processing

These constituted our evaluation metrics, and the statistical formulas employed to calculate them are outlined below.

1. **Accuracy:**
$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$
2. **Loss:**
$$\text{Loss} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$
where N is the number of samples, y_i is the actual value, and $\hat{y}_i$ is the predicted value.
3. **Validation Loss:**
$$\text{Validation Loss} = \frac{1}{M}\sum_{j=1}^{M}(y_j - \hat{y}_j)^2$$
where M is the number of validation samples, y_j is the actual value, and $\hat{y}_j$ is the predicted value on the validation set.
4. **Validation Accuracy:**
$$\text{Validation Accuracy} = \frac{\text{Number of Correct Predictions on Validation Set}}{\text{Total Number of Predictions on Validation Set}}$$

Figure 4: Evaluation Metrics

V. CONCLUSION

The healthcare analytics breakthrough, particularly in predictive patient case similarity, addresses industry challenges with an intricately designed system. Fueled by advanced machine learning, it responds to escalating patient numbers and emergency risks. The system's core innovation lies in state-of-the-art algorithms, emphasizing data preprocessing for reliable input. This meticulous approach enhances model precision, enabling more accurate predictions. Its seamless integration into hospital systems ensures efficiency in managing extensive patient data, making it a valuable asset for healthcare operations. The system's patient-centric care, ethical considerations, and adaptability contribute to transformative healthcare analytics, promising accurate predictions, personalized treatments, and improved outcomes.

ACKNOWLEDGEMENTS

We express profound gratitude to Dr. K Marimuthu, Professor, School of Computer Science Engineering & Information Science, Presidency University, for his inspirational guidance, invaluable suggestions, and providing us the opportunity to showcase our technical capabilities in completing the project. Our heartfelt thanks extend to our family and friends for their unwavering support and inspiration, integral to the successful realization of this project.

VI. REFERENCES

[1] Anis Sharafoddini, Joel A Dubin and Joon Lee, "Patient Similarity in Prediction Models Based on Health Data", JMIR Med Inform 2017 Jan-Mar, 5(1): e7.

[2] LWC Chan, T Chan, LF Cheng, WS Mak, "Machine Learning of Patient Similarity", Bioinformatics and Biomedicines Workshops, 2010 IEEE International Conference. Gyusoo Kim and Seulgi Lee, "2014 Payment Research", Bank of Korea, Vol. 2015, No. 1, Jan. 2015.

[3] Sharafoddini, A.; Dubin, J.; Lee, J. Patient Similarity in Prediction Models Based on Health Data: A Scoping Review. JMIR Med. Inform. 2017, 5, e7.

[4] Roque, F.; Jensen, P.; Schmock, H.; Dalgaard, M.; Andreatta, M.; Hansen, T.; Søeby, K.; Bredkjær, S.; Juul, A.; Werge, T.; et al. Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. PLoS Comput. Biol. 2011, 7, e1002141.

[5] Planet, C.; Gevaert, TO. CoINcIDE: A framework for discovery of patient subtypes across multiple datasets. Genome Med. 2016, 8, 27.

[6] Zhan, M.; Cao, S.; Qian, B.; Chang, S.; Wei, J. Low-Rank Sparse Feature Selection for Patient Similarity Learning. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016.

[7] Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep. 2016;6(1):1–10.

[8] Choi E, Schuetz A, Stewart WF, Sun J. Medical concept representation learning from electronic health records and its application on heart failure prediction. arXiv Prepr arXiv 160203686. 2016;

[9] Savolainen, M. J., Karlsson, A., Ronkainen, SToppila, I., Lassenius, M. I., Falconi, C. V., et al. (2021). The Gaucher Earlier Diagnosis Consensus point-scoring System (GED-C PSS): Evaluation of a Prototype in Finnish Gaucher Disease Patients and Feasibility of Screening Retrospective Electronic Health Record Data for the Recognition of Potential Undiagnosed Patients in Finland. Mol. Genet. Metab. Rep. 27, 100725. doi:10.1016/j.ymgme.2021.1 00725

[10] Barco, T. L., Kuchenbuch, M., Garcelon, N., Neuraz, A., and Nabbout, R. (2021). Improving Early Diagnosis of Rare Diseases Using Natural Language Processing in Unstructured Medical Records: an Illustration from Dravet Syndrome. Orphanet J. Rare Dis. 16, 309. doi:10.1186/s13023-021-01936-9