##### **Accurate and Efficient Prescription Analysis with NLP and OCR**

**Avishikta Bhattacharjee**

School of Computer Science Engineering ,KIIT,

[2205195@kiit.ac.in](mailto:21052780@kiit.ac.in)

***Abstract*-**

Incorporating Optical Character Recognition (OCR) technology and Random Forest classification, this project enhances healthcare by accurately assessing prescription quality and providing personalized medication recommendations.Extracted data is then highlighted for quick comprehension, aiding both doctors and non-professionals. Furthermore, the system attempts to predict the underlying disease based on mentioned symptoms, supporting healthcare professionals. Finally, the project explores the potential for digital prescription storage to improve record-keeping in hospitals. This NLP and OCR powered approach has the potential to revolutionize prescription analysis, enhancing patient care and healthcare efficiency.

1. **Introduction**

The cornerstone of effective healthcare lies in accurate diagnosis and treatment. Prescriptions, outlining the medication a patient should receive, play a crucial role in this process. However, traditional paper-based prescriptions can be time-consuming to decipher, prone to errors in interpretation, and difficult to store and manage electronically. This project tackles these challenges by proposing a novel approach to prescription analysis, powered by the combined strengths of Natural Language Processing (NLP) and Optical Character Recognition (OCR). NLP, a branch of Artificial Intelligence, allows computers to understand and process human language. In this context, NLP will be employed to extract key information from the prescription text, such as medication names, dosages, and instructions.

OCR technology bridges the gap between physical documents and digital information. By integrating OCR, the project will be able to convert text on prescriptions into a machine-readable format. This functionality removes the need for manual data entry, improving accuracy and efficiency. With the extracted information, the system will then highlight crucial aspects of the prescription for rapid comprehension. This not only benefits doctors by reducing interpretation time, but also empowers any authorized individual to understand the prescribed medication at a glance.Furthermore, the project delves into the realm of predictive disease analysis. By leveraging the extracted symptoms mentioned in the prescription, the system attempts to predict the underlying condition the medication is intended to treat. This prediction serves as an additional layer of support for healthcare professionals, potentially aiding in diagnosis and treatment planning.Beyond immediate patient care, this technology also holds significant potential for healthcare institutions. Digital storage of prescriptions, facilitated by OCR, enhances record-keeping efficiency and simplifies retrieval of patient information. This streamlines administrative processes and fosters a more digital healthcare ecosystem.

In conclusion, this project presents a novel approach to prescription analysis, harnessing the power of NLP and OCR. By automating information extraction, highlighting key details, and potentially predicting underlying conditions, the system promises to revolutionize the way prescriptions are interpreted and utilized, ultimately contributing to improved patient care and a more efficient healthcare system**.**

1. **Study of similar projects or technology\ literature review**

Several studies, have explored predictive modeling techniques to assess prescription accuracy. These models, often based on advanced algorithms like Random Forest and Neural Networks, showcase promising results in evaluating the effectiveness of prescribed medications. Additionally, research efforts like have delved into natural language processing, enabling intelligent systems to recommend medications based on patient symptoms and historical data. These studies underline the significance of data-driven decision-making in healthcare.

Moreover, the emergence of electronic health records (EHRs) has revolutionized data accessibility. Integrating EHRs with machine learning algorithms has facilitated comprehensive patient profiling. By incorporating demographic details, medical histories, and real-time symptom data, these projects enhance the accuracy of prescription evaluations and recommendations.

Finally, we use Flask library to provide a web app interface to the project.Flask is a Python framework that simplifies web application development. Unlike bulkier options, Flask offers a lightweight core, allowing you to build web apps tailored to your needs. It handles directing user requests to the appropriate code sections and integrates with templating engines for dynamic web page creation.

1. **Basic concepts/ Technology used**

In this project, Optical Character Recognition (OCR) technology was harnessed to extract textual data from medical prescriptions, ensuring seamless digitization of handwritten or printed records. The system utilized OCR-extracted data to create relevant features, such as patient demographics and symptoms, for the Random Forest Classifier.

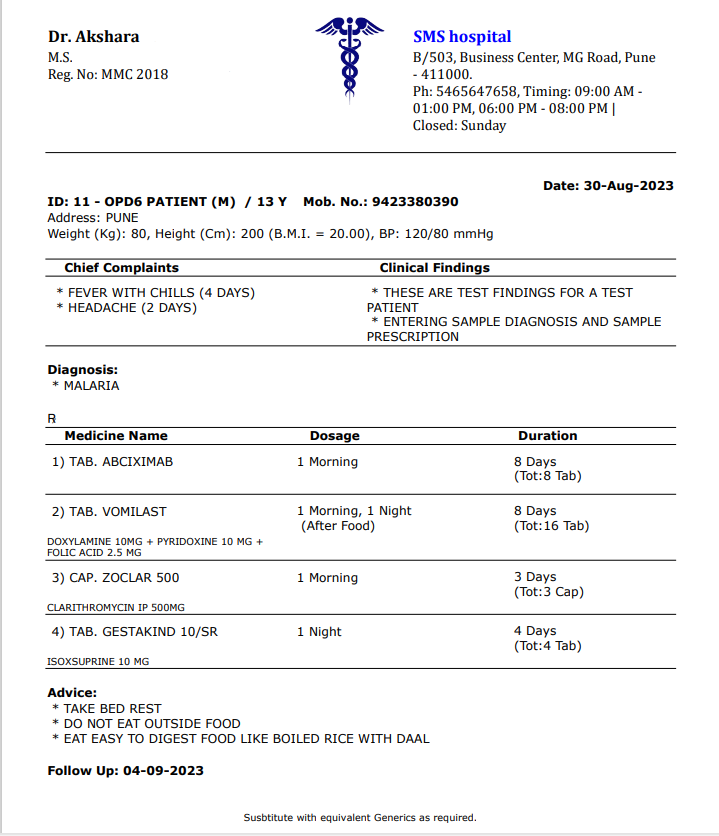
Random Forest, an ensemble learning algorithm, offers significant benefits in diverse fields, especially healthcare. Its ability to handle complex datasets, manage missing values, and mitigate overfitting ensures accurate predictions. In healthcare, it aids in precise diagnosis, drug discovery, and patient outcome predictions, enhancing medical decision-making and ultimately improving healthcare efficacy and efficiency.

1. **Implementation and results**

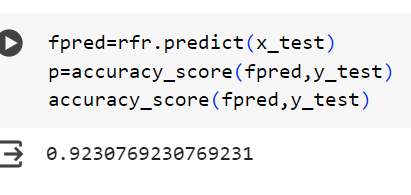
The cornerstone of the project lies in data preparation. Categorical variables, such as symptoms and pre-existing conditions, are converted into numerical labels using a label encoder. This allows the machine learning model to understand and utilize the data effectively. Next, the data undergoes a split using *train\_test\_split*, creating separate training and testing sets. The training set is used to build the model, while the testing set evaluates its performance on unseen data. Within the data, the target variable, "Disease," is identified and assigned to a variable for the model to predict. The remaining features, encompassing vital information like symptoms (fever, cough) and patient age, are also assigned to a separate variable for model training.

The chosen model for this project is a Random Forest, a popular algorithm well-suited for supervised learning tasks with pre-existing datasets. Random Forests excel at handling complex relationships between features and the target variable. The model is trained on the prepared training data, and its performance is evaluated using metrics like accuracy. In this instance, the model achieved an accuracy of 92.3%, demonstrating its effectiveness in predicting disease based on the provided features. However, the project acknowledges the potential for further enhancing accuracy through techniques like boosting algorithms. Finally, the trained model is saved using joblib, allowing it to be loaded and utilized for future predictions.

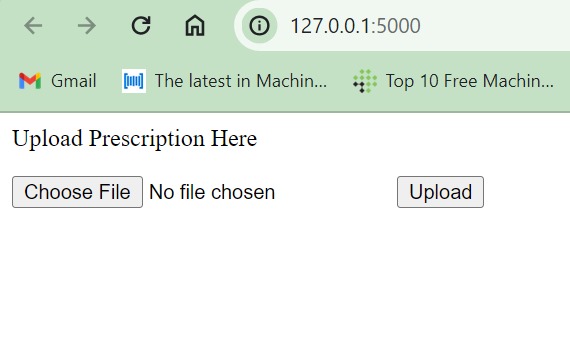
Sample Prescription:



## Model Results:



Web Page:



**SOCIETAL IMPACT AND FUTURE SCOPE**

The integration of machine learning in healthcare, as demonstrated by this project, holds profound societal implications. By enhancing prescription accuracy and providing personalized medication recommendations, this technology ensures safer, more effective treatments. This advancement reduces instances of adverse drug reactions, hospital readmissions, and healthcare costs. Particularly, it empowers individuals in underserved communities, ensuring they receive optimized medical care regardless of geographical constraints. Moreover, the system aids healthcare professionals, enabling them to make data-driven decisions, thereby improving patient trust and healthcare quality.

**1. Integration of Large Language Models (LLMs) for Doctor and Medicine Recommendations:**

**a. LLM Integration:** Incorporate an LLM trained on vast medical data.

**b. Symptom Analysis:** The LLM can analyze the user's symptoms and the predicted disease from the model.

**c. Doctor Recommendation:** Based on the analysis, the LLM can recommend specialists appropriate for the predicted disease, potentially considering factors like location and user preferences.

d. **Medicine Recommendation:** The LLM can also suggest potential medications based on the predicted disease and user's medical history (with appropriate privacy considerations). However, it's crucial to emphasize that these are only suggestions, and a doctor's consultation remains irreplaceable for definitive diagnosis and treatment plans.

2. **Deepening Text Analysis with Attention-based Neural Networks (ANNs) for Semantic Analysis:**

**a. Grid Search with Pipeline:** Utilize a GridSearchCV pipeline to optimize hyperparameters for both the current model and the newly introduced ANN.

**b. Attention-based ANN:** Implement an ANN with an attention mechanism. This allows the model to focus on crucial parts of the prescription text during semantic analysis, leading to a more nuanced understanding.

**c. Semantic Analysis:** The ANN can perform a deeper semantic analysis of the prescription text, potentially extracting medication dosages, durations, and potential side effects.

3. **Convolutional Neural Networks (CNNs) for Handwritten Prescription Processing:**

**a.CNN Integration:** Introduce a CNN specifically trained on handwritten prescription images.

**b.Image Preprocessing:** Develop an image preprocessing pipeline to handle variations in lighting, handwriting styles, and image noise.

**c.Improved Accuracy:** By integrating a CNN, the project can handle handwritten prescriptions seamlessly, improving its overall applicability.

1. **CONCLUSION**

These results underscore the efficacy of integrating machine learning and natural language processing in healthcare contexts. The implemented model and recommendation system not only enhance prescription accuracy but also streamline the decision-making process for healthcare providers, ultimately leading to improved patient outcomes and a more efficient healthcare system.

REFERENCES

1. [Improving Optical Character Recognition of Finnish Historical Newspapers with a Combination of Fraktur & Antiqua Models and Image Preprocessing](https://aclanthology.org/W17-0238) (Koistinen et al., NoDaLiDa 2017)
2. [The Postprocessing of Optical Character Recognition Based on Statistical Noisy Channel and Language Model](https://aclanthology.org/Y95-1017) (Chang & Chen, PACLIC 1995)
3. Ekaterina V. Karmanova, Irina V. Gavrilova, Olga E. Maslennikova, "Deep Learning in Automation of Checking Homework Assignments", 2023 International Russian Smart Industry Conference (SmartIndustryCon), pp.207-212, 2023.
4. J. Jebadurai, I. J. Jebadurai, G. J. L. Paulraj and S. V. Vangeepuram, "Handwritten Text Recognition and Conversion Using Convolutional Neural Network (CNN) Based Deep Learning Model," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1037-1042, doi: 10.1109/ICIRCA51532.2021.9544513. keywords: {Training;Deep learning;Handwriting recognition;Recurrent neural networks;Text recognition;Databases;Training data;Character Recognition;Deep Learning;CTC loss;LSTM;CNN;RNN;Deep Neural Networks},
5. [Distinguishing Clinical Sentiment: The Importance of Domain Adaptation in Psychiatric Patient Health Records](https://aclanthology.org/W19-1915) (Holderness et al., ClinicalNLP 2019)
6. Hercules Dalianis. 2018. Clinical Text Mining: Secondary Use of Electronic Patient Records. Springer International Publishing.
7. Surabhi Adhikari, Surendrabikram Thapa, Priyanka Singh, Huan Huo, Gnana Bharathy, and Mukesh Prasad. 2021. A comparative study of machine learning and nlp techniques for uses of stop words by patients in diagnosis of alzheimer’s disease. In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE.