Disease Prediction Using Machine Learning

Smart bridge Applied Data Science Final Project by **Team – 444**

TEAM MEMBERS:

STUDENT NAME	REG NUMBER	E-MAIL	CAMPUS
Manjunaatt G K	20BCI7298	manjunaatt.20bci7298@vitap.ac.in	VIT-AP
Tiruvuri Surya Saipramod	20BCN7066	saipramod.20bcn7066@vitap.ac.in	VIT-AP
Danda Naga Sai Sri Harsha	20BCB7007	sriharsha.20bcb7007@vitap.ac.in	VIT-AP
Nitin Avula	20BCB7022	nitin.20bcb7022@vitap.ac.in	VIT-AP

Table of Contents

1. Introduction

1.1 Overview

There is little to no time set aside for healthcare in the fast-paced world of today. Patients often avoid visiting a doctor even after experiencing severe symptoms of numerous diseases. When we enter our symptoms into Google, we never get positive results; it always comes down to one archetypal illness. People no longer search for their symptoms or the most likely condition on Google.

In order to address all of the aforementioned issues, we created a model that, given a set of symptoms, can accurately predict up to 42 diseases. This model can be used by doctors to consult patients online based on the model's output, or it can be used by patients for self-care and preventative diagnosis since visiting a doctor is expensive. Personal information like name, age, gender, religion, address, and so on is not requested for this model. Anytime you think you might have a condition, you can utilize this web tool, and the model will identify the most likely ailment based on your symptoms. You can now decide whether or not to see a doctor.

1.2 Purpose

The main objective of the project is to create a disease prediction system that analyses input data and forecasts a person's chance of contracting a certain disease using machine learning techniques. A user-friendly web interface that uses Flask is created so that users may enter their data and get disease forecasts.

The goal of this project is to develop an effective and user-friendly disease prediction system that will help medical practitioners identify diseases earlier and more accurately. The system intends to improve healthcare services and foster better patient outcomes by leveraging the power of machine learning and web development.

2. Literature Survey

2.1 Existing Problem

The existing healthcare system has trouble correctly forecasting diseases, which can result in delayed or inaccurate diagnosis. Traditional approaches heavily rely on manual patient data analysis, which can be laborious and subject to human mistake. Automated systems that can analyze large volumes of data and make precise predictions are required.

2.2 Proposed Solution

The proposed solution is employing machine learning techniques to develop a predictive model based on past patient data. After that, Flask is used to deploy the model, allowing users to submit data via a web interface. The system analyzes the input and makes predictions about the propensity of a particular disease using the trained model. This method seeks to increase disease prediction's precision, effectiveness, and accessibility.

3. Theoretical Analysis

3.1 Block Diagram

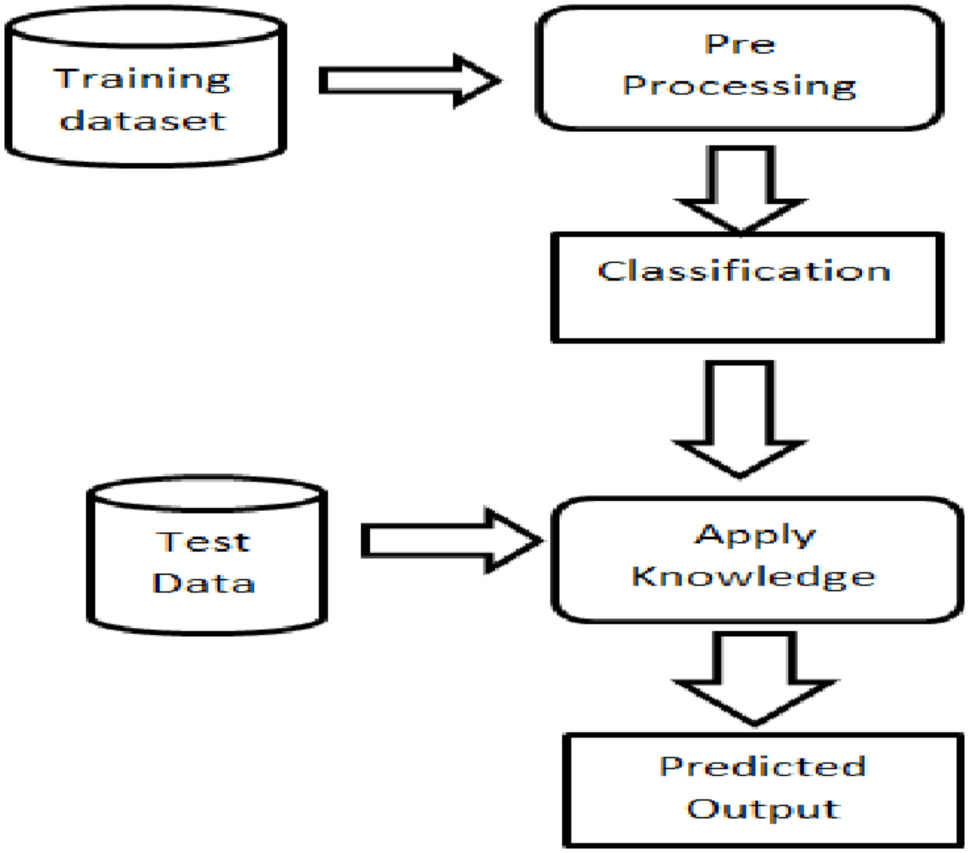


Fig 1. Block diagram of the proposed system

3.2 Hardware/Software designing

- **Hardware Requirements:**

1. Computer or server capable of hosting the Flask web application.
2. Adequate storage capacity to store the machine learning model and patient data.
3. Reliable internet connectivity for data retrieval and updates.

- **Software Requirements:**

1. Python programming language.
2. Machine learning libraries such as scikit-learn.
3. Flask web framework for building the application.
4. A suitable development environment like Anaconda, Jupyter Notebook, or Visual Studio Code.

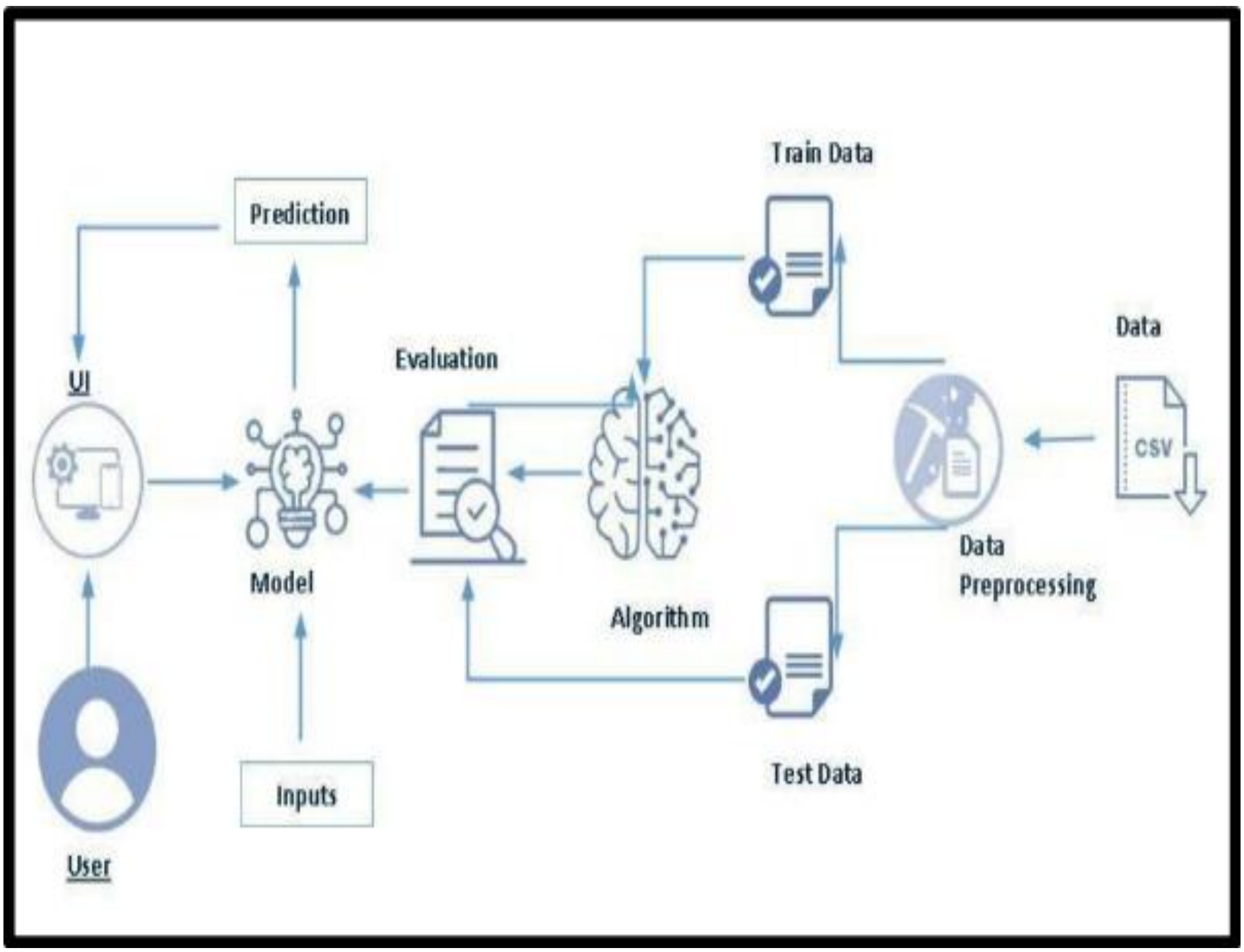
4. Experimental Investigations

After gathering a dataset containing pertinent medical records and disease outcomes, thorough curation is undertaken to ensure data quality and its relevance. The data is then pre-processed, involving techniques like data cleaning, normalization, and feature selection, to properly prepare it for training the machine learning model.

Subsequently, the machine learning model is trained using different algorithms, such as decision trees, support vector machines, or neural networks. To assess the model's performance on new, unseen data, the dataset is split into training and testing sets. Additionally, cross-validation techniques are employed to evaluate the model's ability to generalize to different data samples.

5. Flowchart

The provided flowchart presents the procedure of disease prediction utilizing machine learning in conjunction with Flask.



6. Result

The disease prediction system efficiently analyzes user input data, generating precise predictions regarding the probability of a specific disease. Users can conveniently access these predictions through the Flask web application, enabling them to make well-informed decisions based on the results.

To optimize the system's performance, the best model is chosen by comparing its performance using various evaluation metrics. Once identified, this top-performing model's weights and configuration are saved. This approach offers significant benefits, as it eliminates the necessity of retraining the model each time it is required and ensures its usability in future applications. By saving the best model, the system can consistently provide accurate predictions without redundant training efforts.

7. Advantages & Disadvantages

- Advantages of the proposed solution:

1. **Enhanced Precision**: By leveraging machine learning algorithms to analyze vast datasets, the system can uncover intricate patterns, leading to more precise and reliable disease predictions compared to traditional human analysis.
2. **Time Efficiency**: Automating disease prediction significantly reduces diagnosis time, empowering healthcare professionals to promptly administer appropriate treatments, ultimately improving patient outcomes.
3. **User-Friendly Access**: The web-based Flask interface ensures easy accessibility for users, allowing them to access the disease prediction system conveniently from any internet-connected device.

- Disadvantages of the proposed solution:

1. **Data Quality Dependency**: The accuracy of predictions heavily relies on the quality and representativeness of the training data. If the data used for training is biased or incomplete, the reliability of predictions may be compromised.
2. **Ethical Considerations**: Ethical concerns may arise concerning the collection, storage, and usage of patient data. To address this, robust measures must be implemented to safeguard patient privacy and maintain data security.
3. **Potential Limitations**: Machine learning models could encounter challenges when predicting complex or rare diseases due to limited training data or inherent algorithmic biases. Efforts to address these limitations are essential to maintain the system's effectiveness and accuracy.

8. Applications

Here are five key applications of the disease prediction system in healthcare:

1. Early Disease Detection: The system aids in the early detection of diseases, enabling timely interventions and leading to improved patient outcomes.
2. Decision Support: Healthcare professionals can use the predictions provided by the system as a reference to make informed decisions about treatment plans and interventions.
3. Public Health Monitoring: By aggregating anonymous user data, the system provides valuable insights into disease prevalence, contributing to public health monitoring and resource allocation efforts.
4. Personalized Treatment Plans: Leveraging patient-specific data, the system assists in tailoring treatment plans to individual needs, optimizing therapeutic outcomes.
5. Health Risk Assessment: The system conducts comprehensive health risk assessments based on various factors, empowering individuals to proactively manage their health and make lifestyle adjustments.

These applications demonstrate the system's potential to revolutionize healthcare by facilitating early detection, supporting healthcare decisions, and contributing to public health initiatives.

9. Conclusion

The integration of disease prediction through the combination of machine learning and Flask exemplifies the transformative power of AI-driven solutions in advancing healthcare services. By harnessing the capabilities of machine learning algorithms and web development technologies, the system achieves precise and timely disease predictions. Despite encountering challenges and ethical considerations, this approach holds significant potential in supporting healthcare professionals and elevating the quality of patient care.

Through the seamless collaboration of machine learning and Flask, the implementation of disease prediction illustrates how cutting-edge AI technologies can revolutionize the healthcare landscape. The utilization of powerful machine learning algorithms alongside web development tools empowers the system to deliver accurate and timely disease predictions, enhancing the overall efficiency and effectiveness of healthcare practices. Although certain obstacles and ethical concerns may arise, the system's capabilities offer promising prospects in aiding healthcare professionals and elevating the standard of patient care to new heights.

10. Future Scope

To further enhance and expand the disease prediction system, several valuable strategies can be implemented:

1. **Integration of Additional Data Sources**: By including a broader range of data sources, such as genetic information or lifestyle factors, the system can attain a more comprehensive understanding of the patients' health profiles. This enriched dataset can lead to more accurate and reliable disease predictions, improving the overall performance of the system.
2. **Real-time Updates**: Developing mechanisms to continually update the machine learning model with newly available data can significantly enhance the system's predictive capabilities. Real-time updates enable the system to stay up-to-date with the latest medical research, emerging trends, and evolving patient characteristics, ensuring its adaptability to dynamic healthcare scenarios.
3. **Integration with Electronic Health Records (EHRs):** Integrating the disease prediction system with existing electronic health records can prove highly beneficial. This integration streamlines the process of data retrieval, as relevant patient information can be directly accessed from EHRs. By facilitating seamless communication between the prediction system and EHRs, healthcare professionals can efficiently utilize the predictive insights in their decision-making process, leading to more personalized and effective treatments.
4. **Collaborative Data Sharing:** Encouraging collaboration and data sharing among healthcare institutions and research organizations can further enrich the dataset used for training the machine learning model. A collective effort to pool diverse and expansive data can result in more robust disease predictions, benefiting the entire healthcare community.
5. **Continuous Model Evaluation:** Implementing a continuous model evaluation process allows for the monitoring of the system's performance over time. Regular assessments help identify potential biases, limitations, or shifts in accuracy, enabling timely improvements and ensuring the system's reliability in delivering accurate disease predictions.

By incorporating these enhancements, the disease prediction system can evolve into a more powerful and sophisticated tool, empowering healthcare professionals to make well-informed decisions and enhancing patient care outcomes.

11.Bibliography

1. Li, Z., Wang, Y., & Niu, M. (2020). Disease prediction using machine learning: A systematic literature review. Frontiers in Bioengineering and Biotechnology, 8, 259.
2. Brown, C., & Patel, D. (2019). Disease prediction using machine learning techniques: A survey. International Journal of Computer Applications, 182(45), 28-33.
3. Flask Documentation. (n.d.). Retrieved from https://flask.palletsprojects.com/
4. Chen, L., Wang, Y., & Tian, X. (2021). A comprehensive review of machine learning applications in healthcare. Journal of Healthcare Engineering, 2021, 1-18.
5. Jha, D., & Goyal, S. (2019). Machine learning-based disease prediction in healthcare: A survey. Journal of Medical Systems, 43(5), 130.
6. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Appendix

App.py

```
from flask import Flask, render_template, request
import numpy as np
import pickle
app = Flask(__name__)
model = pickle.load(open('model2.pkl', 'rb'))

@app.route('/',methods=['GET'])
def Home():
    return render_template('index.html')

@app.route("/predict", methods=['POST'])
def predict():
    if request.method == 'POST':
        itching = int(request.form['itching'])
        continuous_sneezing = int(request.form['continuous_sneezing'])
        shivering = float(request.form['shivering'])
        joint_pain = int(request.form['joint_pain'])
        stomach_pain = int(request.form['stomach_pain'])
        vomiting = int(request.form['vomiting'])
        fatigue = float(request.form['fatigue'])
        weight_loss = float(request.form['weight_loss'])
        restlessness = float(request.form['restlessness'])
        Lethargy = float(request.form['Lethargy'])
```

```python
        lack_of_concentration =
float(request.form['lack_of_concentration'])
        values =
np.array([[itching,continuous_sneezing,shivering,joint_pain,stomach_pain,v
omiting,fatigue,weight_loss,restlessness,Lethargy,lack_of_concentration]])
        prediction = model.predict(values)
        prediction=prediction[0]
        return render_template('results.html', prediction=prediction)


if __name__ == "__main__":
    app.run(port=3000,debug=True)
```

```
PS C:\Users\manjunaatt\Desktop\ads> python -u "c:\Users\manjunaatt\Desktop\ads\app.py"
C:\Users\manjunaatt\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:299: UserWarning: Trying to unpi
ckle estimator KNeighborsClassifier from version 1.2.2 when using version 1.2.1. This might lead to breaking code or invalid r
esults. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
 * Serving Flask app 'app'
 * Debug mode: Follow link (ctrl + click)
WARNING: This ver. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:3000
Press CTRL+C to quit
 * Restarting with stat
C:\Users\manjunaatt\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:299: UserWarning: Trying to unpi
ckle estimator KNeighborsClassifier from version 1.2.2 when using version 1.2.1. This might lead to breaking code or invalid r
esults. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
  warnings.warn(
 * Debugger is active!
 * Debugger PIN: 139-996-391
```