CAPSTONE PROJECT

ADVANCED MACHINE LEARNING MODELS FOR EARLY DETECTION OF CORONARY HEART DISEASE

Presented By:
1. SALAM KARISHMA- KONERU LAKSHMAIAH EDUCATION FOUNDATION –Computer Science

edunet foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

- Coronary artery disease (CAD) is a significant global health concern, being one of the leading causes of morbidity and mortality. The complexity of CAD, characterized by the gradual buildup of plaque in the coronary arteries, necessitates early diagnosis and effective risk prediction to improve patient outcomes. Traditional diagnostic methods often rely on clinical assessments and standard tests, which may not fully capture the intricate patterns and interactions present in patient data. As a result, there is a pressing need for advanced predictive models that leverage machine learning and deep learning techniques to analyze large and diverse datasets. These models can provide more accurate risk assessments by identifying subtle correlations and risk factors that may be overlooked by conventional approaches. Addressing this gap in predictive capabilities is crucial for enhancing early detection, guiding treatment decisions, and ultimately reducing the burden of coronary artery disease on individuals and healthcare systems.

PROPOSED SOLUTION

The proposed system aims to develop a robust predictive model utilizing advanced machine learning and deep learning techniques to accurately assess the risk of coronary artery disease (CAD) in patients. The system will leverage the Cleveland Heart Disease dataset, which contains a comprehensive set of clinical and demographic features relevant to heart disease.

Key Components of the Proposed System:

1.Data Collection and Preprocessing:

1. The Cleveland Heart Disease dataset will be utilized, comprising various attributes such as age, sex, blood pressure, cholesterol levels, and other clinical indicators.
2. Data preprocessing steps will include:
    1. **Data Cleaning:** Handling missing values and outliers to ensure data integrity.
    2. **Normalization:** Scaling numerical features to a standard range to improve model performance.
    3. **Encoding:** Converting categorical variables into numerical formats using techniques like one-hot encoding or label encoding.

2.Feature Selection:

1. Employing feature selection techniques to identify the most significant predictors of CAD. This step is crucial for reducing dimensionality and enhancing model interpretability.
2. Techniques such as correlation analysis and recursive feature elimination will be used to determine the relevance of each feature.

3. Model Development:

1. The core of the proposed system will be the implementation of Artificial Neural Networks (ANN). This model is chosen for its ability to capture complex, non-linear relationships within the data.
2. The ANN architecture will be designed with multiple layers, including input, hidden, and output layers, to facilitate deep learning capabilities.

4. Training and Validation:

1. The model will be trained using a portion of the dataset, with the remaining data reserved for validation and testing.
2. Cross-validation techniques will be employed to ensure the model's robustness and to prevent overfitting.

5. Risk Prediction:

1. Once trained, the model will be capable of predicting the likelihood of CAD in new patients based on their clinical data.
2. The output will be a probability score indicating the risk level, which can assist healthcare professionals in making informed decisions regarding patient management.

6. User Interface:

1. A user-friendly interface will be developed to allow healthcare providers to input patient data easily and receive risk predictions.
2. The interface will also provide visualizations of the results, such as risk factor contributions and model confidence levels.

7. Integration and Deployment:

1. The final model will be deployed as a web application, enabling easy access for healthcare professionals.
2. Continuous updates and improvements will be made based on user feedback and new data, ensuring the model remains relevant and accurate over time.

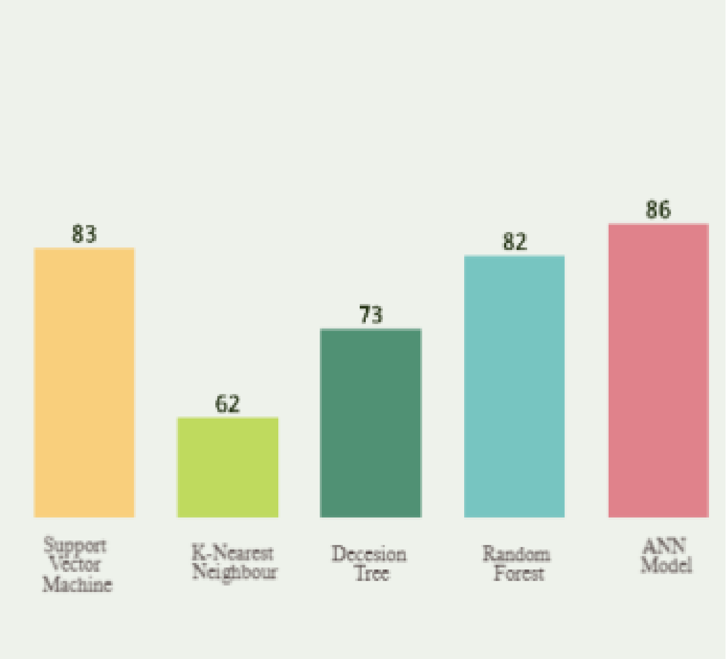
SYSTEM APPROACH

The "System Approach" for developing the coronary artery disease (CAD) prediction system includes hardware with a minimum of 8 GB RAM and 10 GB storage, software requirements of Python 3.7 or higher with an IDE like Jupyter Notebook, and essential libraries such as pandas, numpy, matplotlib, seaborn, scikit-learn, and tensorflow or keras for data manipulation, visualization, machine learning, and model evaluation.

ALGORITHM & DEPLOYMENT

- **Algorithm:**
  - Artificial Neural Networks (ANN) for modeling complex relationships in the data.
  - Feature selection and correlation analysis to identify significant predictors.
  - Cross-validation techniques to ensure model robustness.
- **Deployment:**
  - Deployment of the model as a web application using cloud services (e.g., AWS).
  - User interface for healthcare professionals to input patient data and receive risk predictions.

RESULT



Accuracy Comparison of Different Models

	Accuracy	Precision	Recall	F1-Score
SVM	83.33	70	95.45	80.77
KNN	61.67	48.72	86.36	62.3
Decesion Tree	73.33	60	81.81	69.23
Random Forest	81.67	67.74	95.45	79.25
ANN	84.44	80	84.21	82.05

Perfomance metrics of Models

CONCLUSION

- In conclusion, the development of a coronary artery disease (CAD) prediction system using machine learning techniques offers a promising approach to enhance early diagnosis and risk assessment. By leveraging advanced algorithms such as Artificial Neural Networks and employing comprehensive data preprocessing and evaluation methods, the system aims to provide accurate predictions that can significantly aid healthcare professionals in making informed decisions. This initiative not only contributes to improved patient outcomes but also emphasizes the importance of integrating technology in medical diagnostics.

FUTURE SCOPE

- The future scope of the CAD prediction system includes exploring alternative deep learning architectures, such as AlexNet and RNN, to enhance prediction accuracy. Additionally, there is potential for optimizing the activation functions used in the models. Developing a user-friendly web application to utilize the prediction system with real-time user data is also planned, with a preference for using Amazon Web Services (AWS) for its superior security and range of services compared to other platforms. This will facilitate broader accessibility and practical application of the predictive model in clinical settings.

REFERENCES

Include relevant academic papers, articles, and datasets used in your research.

- Example references:
  - Kim, J. K., & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis.
  - Ayatollahi, H., Gholamhosseini, L., & Salehi, M. (2019). Predicting coronary artery disease: a comparison between two data mining algorithms.
  - Other relevant studies and sources.

COURSE CERTIFICATE 1


In recognition of the commitment to achieve professional excellence

Getting Started with Enterprise-grade AI

IBM SkillsBuild

SALAM KARISHMA

Has successfully satisfied the requirements for:

Getting Started with Enterprise-grade AI

Issued on: 18 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/g9n45fCA

IBM

COURSE CERTIFICATE 2

In recognition of the commitment to achieve professional excellence

Getting Started with Data

IBM SkillsBuild

SALAM KARISHMA

Has successfully satisfied the requirements for:

Getting Started with Data

Issued on: 27 JUL 2024

Issued by IBM

Verify: https://www.credly.com/go/Xbtnoled

IBM

THANK YOU