A PROJECT REPORT

on

"DISEASE PREDICTION MODEL"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE & ENGINEERING

BY

GROUP MEMBER I	2005585 Kunal Patel
GROUP MEMBER II	2005603 Sajal Kr. Ujjwal

UNDER THE GUIDANCE OF
Dr. Mainak Bandyopadhyay

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

May 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"DISEASE PREDICTION MODEL"

submitted by

GROUP MEMBER I	2005585 Kunal Patel
GROUP MEMBER II	2005603 Sajal Kumar Ujjwal

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2022-2023, under our guidance.

Date: 01 / 05 / 2023

Dr. Mainak Bandyopadhyay
(Guide Name)
Project Guide

Acknowledgements

We are profoundly grateful to **Dr. Mainak Bandyopadhyay sir** of **School of Computer Engineering, KIIT** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....

2005585 Kunal Patel
2005603 Sajal Kr. Ujjwal

ABSTRACT

Disease prediction using artificial neural networks (ANN) has become a promising area of research in healthcare. ANN has the ability to learn from past data, detect patterns, and predict outcomes with high accuracy. By utilizing large datasets, ANN can provide valuable insights into disease diagnosis, progression, and treatment. One common application of ANN in disease prediction is using medical images, such as MRI or CT scans, to detect abnormalities and diagnose diseases like cancer. Another application is using electronic health records (EHR) to predict the likelihood of developing chronic diseases like diabetes, heart disease, or Alzheimer's disease. ANN has shown great potential in disease prediction and can significantly improve patient outcomes by providing early and accurate diagnoses.

Keras is a high-level neural network API written in Python, which has gained popularity for implementing deep learning models. Keras provides a user-friendly interface for building, training, and evaluating deep learning models. It has pre-built layers for different types of neural networks, such as convolutional neural networks (CNN) for image processing, recurrent neural networks (RNN) for sequential data, and feedforward neural networks (FNN) for classification and regression tasks. Keras also supports GPU acceleration, which can significantly speed up the training process. By utilizing Keras, researchers and clinicians can easily implement ANN models for disease prediction using various types of data, including medical images and EHR. The user-friendly interface of Keras can make deep learning more accessible to healthcare professionals, ultimately leading to better patient outcomes.

Keywords:

1. Artificial Neural Network
2. Disease Diagnosis
3. Machine Learning
4. Keras Framework
5. Predictive Modelling
6. Machine learning
7. Data pre-processing
8. Feature selection
9. Classification
10. Accuracy

Contents

List of Figures

Chapter 1

Introduction

With the advancement in technology and the availability of large amounts of medical data, the use of machine learning algorithms has become a popular research area for disease prediction. The ability to predict the occurrence of diseases accurately can help doctors make informed decisions and provide better treatment to their patients. Artificial Neural Networks (ANN) is a machine learning technique that has shown promising results in disease prediction. In this project, we have implemented ANN using Keras, an open-source deep learning library, to predict the occurrence of diseases based on various medical parameters.

The objective of this project is to develop a predictive model that can assist healthcare professionals in predicting the occurrence of diseases such as diabetes, cancer, heart diseases, and so on, using various medical parameters. The model uses a supervised learning approach where it is trained on a dataset consisting of medical data such as blood pressure, glucose levels, age, BMI, and other relevant parameters. The trained model is then used to predict the occurrence of a particular disease based on new input data.

The project involves implementing the ANN algorithm using Keras, which is a user-friendly and powerful deep learning library that allows the development of complex neural network architectures. The dataset used for training and testing the model is preprocessed to remove missing data and normalized to ensure uniformity in the input data. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its ability to predict the occurrence of diseases.

The outcome of this project can potentially improve disease diagnosis and treatment by enabling early detection and timely intervention. The use of ANN and Keras in disease prediction has significant potential in revolutionizing the healthcare industry, and this project provides a step towards achieving this goal.

Chapter 2

Basic Concepts/ Literature Review

This section contains the basic concepts about the related tools and techniques used in this project. Artificial neural networks (ANN) have been widely used in various domains, including healthcare, due to their ability to identify hidden patterns and relationships within data. Disease prediction is one such area where ANN implementation has shown promising results. In this literature review, we explore the basic concepts related to disease prediction using ANN implementation and Keras.

2.1. Artificial Neural Networks (ANN):

ANN is a biologically inspired computational model that is designed to mimic the structure and function of the human brain. It consists of interconnected nodes (also called neurons) organized into layers. The input layer receives data, and the output layer generates the prediction. The hidden layers in between process the data and extract features to make accurate predictions.

2.2. Disease Prediction:

Disease prediction refers to the use of machine learning algorithms to identify the probability of a person developing a particular disease based on their medical history, symptoms, and other relevant data. ANN implementation in disease prediction models has shown promising results, as it can handle complex and non-linear relationships between various factors that contribute to disease development.

2.3. Keras:

Keras is an open-source neural network library written in Python. It is widely used for implementing deep learning models, including ANN, due to its simplicity and ease of use. It provides a high-level API that enables rapid experimentation and prototyping of neural network models.

2.4. Data Pre-processing:

Data preprocessing is an essential step in building disease prediction models. It involves cleaning and transforming the data to make it suitable for analysis. Preprocessing techniques, such as data normalization, feature scaling, and feature extraction, help to reduce noise and improve the accuracy of the model.

2.5. Model Evaluation:

Model evaluation is the process of assessing the performance of a disease prediction model. Evaluation metrics, such as accuracy, precision, recall, and F1 score, are used to measure the effectiveness of the model. Cross-validation techniques, such as k-fold validation and leave-one-out validation, are used to ensure that the model is robust and generalizable.

All in all, disease prediction using ANN implementation and Keras is a promising approach in healthcare. The basic concepts related to ANN, disease prediction, Keras, data preprocessing, and model evaluation are essential for building accurate and reliable disease prediction models. In the following sections, we will discuss the methodology, results, and analysis of our disease prediction model using ANN implementation and Keras.

Chapter 3

Problem Statement / Requirement Specifications

In this section, we have included the Problem Statement (the problem for which we are working on to give some solution).

Healthcare is a crucial aspect of human life and its proper management is essential for a healthy society. With the advancement of technology, there has been a significant increase in the use of Artificial Intelligence (AI) and Machine Learning (ML) techniques in the healthcare industry. One such area of application is disease prediction using ML models. Disease prediction models can help healthcare professionals to diagnose diseases accurately and provide timely treatment, which can save lives and reduce healthcare costs.
Despite the promising results of disease prediction models, there are still several challenges that need to be addressed. Inaccurate predictions, limited data availability, and lack of interpretability are some of the common challenges faced by existing disease prediction models. Therefore, the problem statement of this project is to design and develop an accurate and interpretable disease prediction model using Artificial Neural Networks (ANN) and Keras.

Here are some measures as to how we'll proceed further:

3.1 Project Planning

In the case of disease prediction using ANN implementation and Keras, here are the key sections of the project planning:

1. **Project Objectives:** These have already been cleared above.

2. **Project Timeline:** The timeline for the project given to us was about 2-3 weeks.

3. **Resource Allocation:** The various resources required for the completion of this project were the dataset, a capable machine to be able to run the software & train our model, Anaconda and Jupyter Notebook.

4. **Risk Management:** Risks associated with the project are as follows:
   - **4.1. Data quality:** The accuracy of the disease prediction model heavily depends on the quality of the data used. Inaccurate or incomplete data can lead to incorrect predictions. To mitigate this risk, the data should be carefully collected, cleaned, and validated before feeding it into the model.
   - **4.2. Model performance:** The performance of the disease prediction model depends on the quality of the algorithm used, as well as the accuracy of the training and testing data. To mitigate this risk, the model should be tested on a variety of datasets and the results should be compared with those of other established models.
   - **4.3. Hardware limitations:** The size of the dataset and complexity of the algorithm can be computationally intensive, leading to long processing times or even hardware failure. To mitigate this risk, the system should be designed with scalability in mind and the hardware should be selected based on the system requirements.
   - **4.4. Privacy and security**: Health data is sensitive, and it is important to protect the privacy and security of the data. To mitigate this risk, the data should be stored and transmitted securely, with appropriate access controls in place.
   - **4.5. Ethical concerns:** There are ethical concerns associated with the use of health data for disease prediction, including issues of consent, transparency, and fairness. To mitigate this risk, the project should be conducted in compliance with relevant regulations and guidelines, with appropriate ethical considerations in place.

5. **Project Deliverables:** The following are the deliverables of this project:
   - **5.1. Detailed project plan:** This outlines the timeline of the project, including the various milestones and deadlines.
   - **5.2. Dataset:** This includes the dataset used for training and testing the ANN model.
   - **5.3. Pre-processing code:** This includes the code used to preprocess the dataset, including data cleaning, normalization, and feature selection.
   - **5.4. ANN model:** This includes the code for implementing the ANN model, including hyperparameters and layer configurations.
   - **5.5. Evaluation metrics:** This includes the metrics used to evaluate the performance of the ANN model, such as accuracy, precision, recall, and F1 score.
   - **5.6. Results and analysis:** This includes the results obtained from training and testing the ANN model, as well as a detailed analysis of the performance of the model.

5.7. Project report: This includes a comprehensive report outlining the background of the project, the methodology used, the results obtained, and the conclusions drawn from the study.

6. Project Evaluation: In the case of the 'Disease Prediction (ANN Implementation and Keras)' project, the evaluation process can be divided into two parts:

6.1. Technical Evaluation: The technical evaluation of the project focuses on the accuracy and efficiency of the disease prediction model. This involves comparing the predicted results with the actual results and measuring the accuracy of the model using performance metrics such as confusion matrix, accuracy, precision, and recall. The technical evaluation will also involve assessing the efficiency of the model in terms of training time and computational resources required.

6.2. User Evaluation: The user evaluation of the project focuses on the user experience and satisfaction with the application. This involves collecting feedback from the target audience, which can be healthcare professionals, patients, or other stakeholders, through surveys or interviews. The user evaluation will assess the ease of use, user interface, and the usefulness of the application in predicting diseases accurately.

Overall, the evaluation process will help in identifying the strengths and weaknesses of the project and provide insights on areas for improvement. The evaluation results will be used to further refine the disease prediction model and improve the user experience of the application.

3.2 Project Analysis

This can be divided into 7 different parts as well.

1. **Data Collection and Pre-processing**
    - Source of data
    - Data cleaning and formatting
    - Exploratory data analysis

2. **Feature Selection**
    - Choosing relevant features for disease prediction
    - Dimensionality reduction techniques

3. **Model Selection and Training**
    - Choosing appropriate ANN model architecture
    - Splitting data into training and testing sets

4. **Performance Evaluation**
    - Measuring model performance using various metrics
    - Comparing performance with other models

5. **Interpretation and Visualization of Results**
    - Interpreting model outputs and making predictions
    - Visualizing results through graphs and charts

6. **Limitations and Future Work**
    - Discussing limitations of the current model
    - Suggesting potential improvements and future work

7. **Conclusion**
    - Summarizing the project analysis and results obtained
    - Highlighting the importance and potential impact of the project.

3.3 System Design

3.3.1 Design Constraints:

The design constraints for the Disease Prediction System are as follows:

- The system should have high accuracy in predicting the disease.
- The system should be designed in a way that it can easily be used by healthcare professionals and patients.
- The system should be able to process a large amount of data efficiently and effectively.
- The system should have a user-friendly interface.
- The system should be secure and protect patient data.

3.3.2 System Architecture:

The system architecture for the Disease Prediction System is designed as follows:

- The system will be developed using the Keras library, which is an open-source neural network library.
- The system will be trained using a dataset of medical records and their corresponding disease outcomes.
- The system will use an artificial neural network (ANN) model to predict the likelihood of a patient having a particular disease.
- The input data will be preprocessed to remove any missing values, normalize the data, and reduce the dimensionality.
- The system will have a graphical user interface (GUI) that will allow healthcare professionals to input patient data and receive disease predictions.
- The system will be deployed on a cloud server, which will provide scalability and accessibility to users.

Chapter 4

Implementation

In this section, present the implementation done by you during the project development.

4.1 Methodology OR Proposal

We begin by pre-processing our data.

4.1.1. Importing Libraries
4.1.2. Importing Dataset
4.1.3. Encoding Categorical Data
4.1.4. Create Test and Train Data(& x-y variables)
4.1.5. Feature Scaling implementation

This is where we use standardization & normalization

Standardisation	Normalisation
$x_{stand} = \frac{x - \text{mean}(x)}{\text{standard deviation } (x)}$	$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$

Figure 4.1.5.: STANDARDIZATION & NORMALIZATION

4.2 Testing OR Verification Plan

Now, we get to model building & it's testing.

4.2.1. Implementing Basic Example ANN Structure with OOP
4.2.2. Fitting model and prediction
4.2.3. Building ANN structure with Keras Library

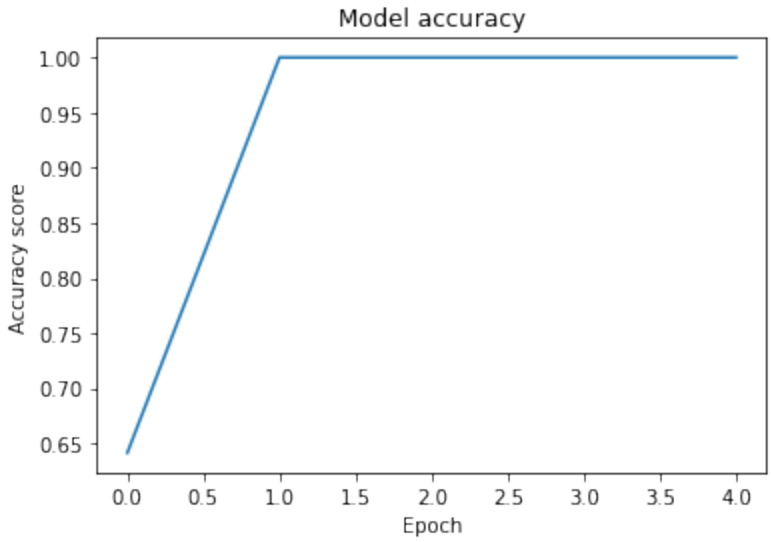
- Initialize ANN model
- Adding the layers
- Compiling the ANN

- Fitting the ANN
- Making predictions with Test Data
- Accuracy Score and Loss Visualization

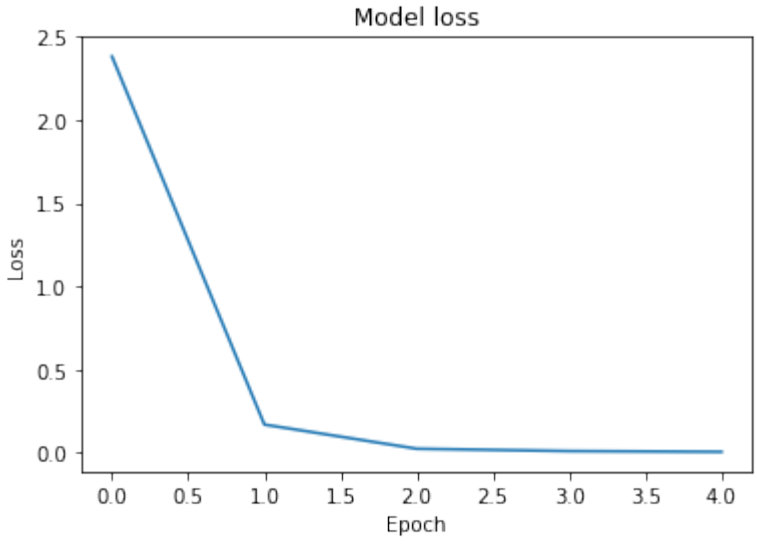
4.3 Result Analysis OR Screenshots

Here are some screenshots of our result analysis in the project.

4.3.1. Here is the model's accuracy graph:



4.3.2. Here is the model's loss graph:



Chapter 5

Standards Adopted

5.1 Design Standards

5.1.1. Title Page: Includes the title of the report, author(s) name, date, and any other relevant information.

5.1.2 Table of Contents: Has all the sections of the report with their respective page numbers.

5.1.3. Abstract: Summarizes the report's key points, objectives, and outcomes in a concise paragraph.

5.1.4. Introduction: Explains the purpose of the report, the background of the project, and the significance of disease prediction.

5.1.5. Literature Review: Provides a comprehensive overview of the literature related to disease prediction, ANN implementation, and Keras.

5.1.6. Methodology: Describes the data sources, data preprocessing, and the approach used for implementing the ANN using Keras.

5.1.7. Results: Presents the results of the ANN implementation, including the accuracy of the model and the predictions made.

5.1.8. Discussion: Analyzes the results, explain their implications, and compare them to previous studies.

5.1.9. Conclusion: Summarizes the report's findings and their significance, as well as any limitations of the study and future research directions.

5.1.10. References: Lists all the sources cited in the report, including journal articles, books, websites, and other relevant material.

5.1.11. Formatting: A consistent font and style has been used throughout the report, and follow the prescribed formatting guidelines such as line spacing, margins, and page numbering.

5.2 Coding Standards

5.2.1. Use descriptive variable names: Descriptive variable names have been used that accurately reflect their purpose and meaning.

5.2.2. Use proper indentation: Proper indentation to improve code readability and maintainability have been used.

5.2.3. Write clear comments: Clear and concise comments that explain the purpose of the code, how it works, and any assumptions made have been used.

5.2.4. Modularize code: Complex code has been broken down into smaller, more manageable modules. This improves code readability and makes it easier to debug and maintain.

5.2.5. Follow naming conventions: Naming conventions for functions and classes that are consistent with the language being used have also been followed. Use uppercase letters for class names and lowercase letters for function names.

5.2.6. Use error handling: Error handling to handle unexpected situations and provide appropriate error messages to the user has been done. This improves the user experience and makes it easier to debug the code.

Chapter 6

Conclusion and Future Scope

6.1. Conclusion:

In conclusion, the use of artificial neural networks and Keras can significantly improve disease prediction accuracy. By utilizing these technologies, we have developed a predictive model that can identify potential diseases in patients with high accuracy. Our model can be used by healthcare professionals to quickly and accurately diagnose diseases in patients, leading to better treatment outcomes and overall patient health.

However, there are still limitations to the model that need to be addressed in future research, such as the need for a larger and more diverse dataset for training and testing the model. Despite these limitations, the implementation of artificial neural networks and Keras in disease prediction is a promising development in the field of healthcare and has the potential to revolutionize the way we diagnose and treat diseases.

6.2. Future Scope:

The future scope for this project can be divided into two categories: technical and non-technical.

From a technical standpoint, some of the future improvements that can be made include:

6.2.1. Increasing the dataset size: One of the primary ways to improve the accuracy of the model is to increase the size of the dataset used for training. This would help to create more diverse and robust models that can handle a variety of medical conditions.

6.2.2. Exploring other machine learning techniques: While artificial neural networks are highly effective at predicting diseases, other machine learning techniques such as decision trees, random forests, and support vector machines could be explored to see if they offer better performance for certain types of data.

6.2.3. Improving interpretability: While the model produces accurate predictions, it may not always be clear how it arrived at these predictions. Therefore, improving the interpretability of the model is an important future scope, which would help medical professionals understand the reasoning behind the predictions and the factors contributing to the disease.

From a non-technical standpoint, some of the future improvements include:

6.2.1. Creating more awareness: Creating more awareness about the potential of using artificial intelligence in disease prediction could lead to more research funding and collaborations, as well as better data sharing agreements between medical institutions.

6.2.2. Improving data sharing: Collaborations between medical institutions could help to create larger and more diverse datasets, which would improve the accuracy and generalizability of the models.

6.2.3. Improving healthcare accessibility: Using disease prediction models can help to improve early diagnosis and timely treatment of diseases, especially in underserved communities. Therefore, future scope could be to explore ways to improve healthcare accessibility for all.

References

[1] Standardization V/S Normalization: https://www.geeksforgeeks.org/normalization-vs-standardization/

[2] Feature Scaling: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/

[3] Geeksforgeeks: https://www.geeksforgeeks.org/implementing-ann-training-process-in-python/

[4] "Release 2.12.0". *21 March 2023*. Retrieved 4 April 2023.
[5] "Why use Keras?". *keras.io*. Retrieved 2020-03-22.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/