**A PROJECT ON**

****

**AI Enabled Heart Disease Diagnosis System**

****

**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER ENGINEERING**

**Submitted by:**

**Abhimanyu Kumar** **2014359**

**Harsh Yadav** **2014397**

***Under the Guidance of***

**Prof. (Dr.) Sachin Sharma**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**MAY-2023**



**CANDIDATE’S DECLARATION**

I/We hereby certify that the work which is being presented in the Project Report entitled **“AI Enabled Heart Disease Diagnosis System”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Engineering and submitted in the Department of Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun is an authentic record of my own work carried out during a period from **August-2022 to May-2023** under the supervision of **Prof. (Dr.) Sachin Sharma** Department of Computer Science and Engineering, Graphic Era (Deemed to be University).The matter presented in this dissertation has not been submitted by me/us for the award of any other degree of this or any other Institute/University.

**Abhimanyu Kumar** **2014359**

**Harsh Yadav** **2014397**

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

**Supervisor** **Head of the Department**

**External Viva**

**Name of the Examiners: Signature with Date**

1.

2.

**Abstract**

This project proposes an artificial intelligence (AI)-driven heart disease prediction model based on the CSO-CLSTM architecture. As part of the model, convolutional and LSTM layers are combined to capture spatial and temporal dependencies within the input data. Our model can be further improved by utilizing Class-Specific Optimization (CSO), which enables the model to learn class-specific features that can be used to predict heart disease. We trained and validated the proposed model using a large dataset of clinical data, including demographic information, medical history, and results of diagnostic tests. We obtained this dataset from a real-world electronic health record (EHR) system which contains patients with a wide range of risk factors and degrees of disease severity. We train the model using a deep learning framework, and we optimize it using state-of-the-art techniques to ensure its accuracy and robustness. A number of parameters, such as precision, specificity, sensitivity, and the region of the ROC, are used to compute the efficiency of the model. The suggested model is superior compared to current heart disease prediction models, according to experimental findings, with over 90% accuracy and excellent precision and sensitivity. In terms of generalization to other patient populations and datasets, the CSO-CLSTM model also exhibits encouraging findings. By assisting healthcare professionals in the early diagnosis and intervention of individuals at risk of heart disease, the AI-enabled heart disease prediction model employing CSO-CLSTM has the potential to profoundly improve clinical practice. This may result in better patient care, lower medical expenses, and better general health outcomes. Future study may include more model validation using bigger datasets and actual clinical situations, as well as model integration into clinical decision support systems for use in real-world healthcare settings.

**Keywords:** Heart disease prediction; Artificial enabled; CSO-CLSTM Model

**Acknowledgement**

Any achievement, be in scholastic or otherwise does not depend solely on the individual effort but on the guidance, encouragement and co-operation of intellectuals, elders and friends. A number of personalities in their own capacity have helped me in carrying out this project work.

Our sincere thanks to project guide **Prof. (Dr.) Sachin Sharma** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), for his valuable guidance and support throughout the course of project work and for being a constant source of inspiration.

We extend our thanks to **Prof. (Dr.) Guru Prasad M.S.**, Project coordinator, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), for his valuable suggestions throughout all the phases of the Project Work.

We are extremely grateful to **Prof. (Dr.) D. P. Singh**, HOD of the Computer Science and Engineering Department, Graphic Era (Deemed to be University), for his moral support and encouragement.

We thank the **management of Graphic Era (Deemed to be University)** for the support throughout the course of our Bachelor’s Degree and for all the facilities they have provided.

Last, but certainly not least we thank all teaching and non-teaching staff of Graphic Era (Deemed to be University) for guiding us in the right path. Most importantly we wish to thank our parents for their support and encouragement.

Abhimanyu Kumar 2014359

Harsh Yadav 2014397

**Table of Contents**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | **Contents** | **Page No.** | | Abstract | i | | Acknowledgement | ii | | Table of Contents | iii | | List of Tables | iv | | List of Figures | v | | **Chapter 1 Introduction** | **1-3** | | 1.1 Project Introduction | 1 | | 1.2 Problem Statement | 2 | | 1.3 Objectives | 3 | | **Chapter 2**  **Literature Survey/ Background** | **4-5** | | **Chapter 3**  **Requirements and Methodology** | **6-11** | | 3.1 Requirements |  | | 3.1.1 Hardware Requirements | 6 | | 3.1.2 Software Requirements | 6 | | 3.2 Methodology  3.2.1 Download the Cleveland heart disease dataset and format it  3.2.2 Develop an AI using CSO-CLSTM model | 7  8 | | **Chapter 4**  **Testing** | **13** | | **Chapter 5 Conclusion and Future Work** | **14** | | **Details of Research Publication** | **15** | | **References** |  | |  |
|  | **List of Tables**   |  |  |  | | --- | --- | --- | | **TABLE No.** | **TITLE** | **PAGE No.** | | 5.1 | Details of Research Publication | 15 |   **List of Figures**   |  |  |  | | --- | --- | --- | | **FIGURE No.** | **TITLE** | **PAGE No.** | | 3.1 | Flowchart of the Project | 6 | | 3.1 | Flowchart of CSO Algorithm | 9 | | 3.2 | Structure of CLSTM | 10 | | 4.1 | shows prediction on a sample data | 13 | |  |

**Chapter 1**

**Introduction**

In the following sections, a brief introduction and the problem statement for the work has been included.

* 1. **Introduction**

The recent advancements in Internet of Things (IoT), cloud computing, and Artificial Intelligence (AI) transformed the conventional healthcare system into smart healthcare. By incorporating key technologies such as IoT and AI, medical services can be improved. The convergence of IoT and AI offers different opportunities in healthcare sector. In this view, the project presents a new AI and IoT convergence-based disease diagnosis model for smart healthcare system. The major goal of this project is to design a disease diagnosis model for heart disease using AI and IoT convergence techniques. The presented model encompasses different stages namely, data acquisition, pre-processing, classification, and parameter tuning. IoT devices such as wearables and sensors permit seamless data collection while AI techniques utilize the data in disease diagnosis. The proposed method uses Crow Search Optimization algorithm-based Cascaded Long Short-Term Memory (CSO-CLSTM) model for disease diagnosis. In order to achieve better classification of the medical data, CSO is applied to tune both ‘weights’ and ‘bias’ parameters of CLSTM model. The application of CSO helps in considerable improvement in the diagnostic outcomes of CLSTM model. The performance of CSO-LSTM model was validated using healthcare data. During the experimentation, the presented CSO-LSTM model accomplished the maximum accuracies of 96.16% and 97.26% in diagnosing heart disease and diabetes respectively. Therefore, the proposed CSO-LSTM model can be employed as an appropriate disease diagnosis tool for smart healthcare systems.

AI diagnosis has exceeded the accuracy accomplished manually. In addition, ML-based models are precise in comparison with well-trained physicians, particularly pathologists and imaging experts.

* 1. **Problem Statement**

The aim is to develop a disease diagnosis model using AI and IoT convergence technique for the diagnosis of diabetes and heart disease

* Designing and development of a novel AI enabled disease diagnosis for smart healthcare system
* Using CSO-CLSTM model for diagnosing heart disease
* Performed parameter tuning of LSTM model using CSO algorithm
* Validated the performance of CSO-LSTM model on datasets.
  1. **Objectives**

1. Download Cleveland heart disease dataset and format it.
2. To Design AI enabled heart disease diagnosis using CSO-CLSTM model.
3. Validate the performance of CSO-CLSTM model on datasets.

**Chapter 2**

**Literature Survey/ Background**

Predicting heart disease has attracted a lot of scientific attention, and several studies have used artificial intelligence (AI) approaches to create predictive models. The application of CNNs for heart disease prediction has been investigated in several research, with a particular emphasis on the extraction of spatial information from medical imaging data, such as echocardiograms or angiograms. For instance, a CNN-based model was created by Attia et al. [2] (2017) to forecast cardiovascular risk using echocardiography pictures, and it was 80% accurate in foretelling future cardiovascular events. Similar to this, Margetaki et al. [3] (2020) suggested a CNN-based model to forecast severe adverse cardiovascular events using coronary angiography pictures, reaching an 89% accuracy rate.

However, temporal relationships in sequential data, such as time-series data from electronic health records (EHRs), have been successfully captured using RNNs, notably LSTM. For instance, Choi et al. [4] (2016) reported an LSTM-based model with an accuracy of 80% for predicting heart failure using EHR data. An LSTM-based model was created by Li et al. [5] (2019) with an area under the ROC curve of 0.88 to predict long-term mortality in patients with heart failure. In addition, current research has looked into hybrid models that combine CNNs and RNNs to capture both spatial and temporal characteristics for heart disease prediction. As an illustration, Kim et al. [6] (2019) suggested a hybrid CNN-LSTM model for predicting cardiovascular events utilising both EHR data and medical pictures, and it was successful in 92% of the cases.

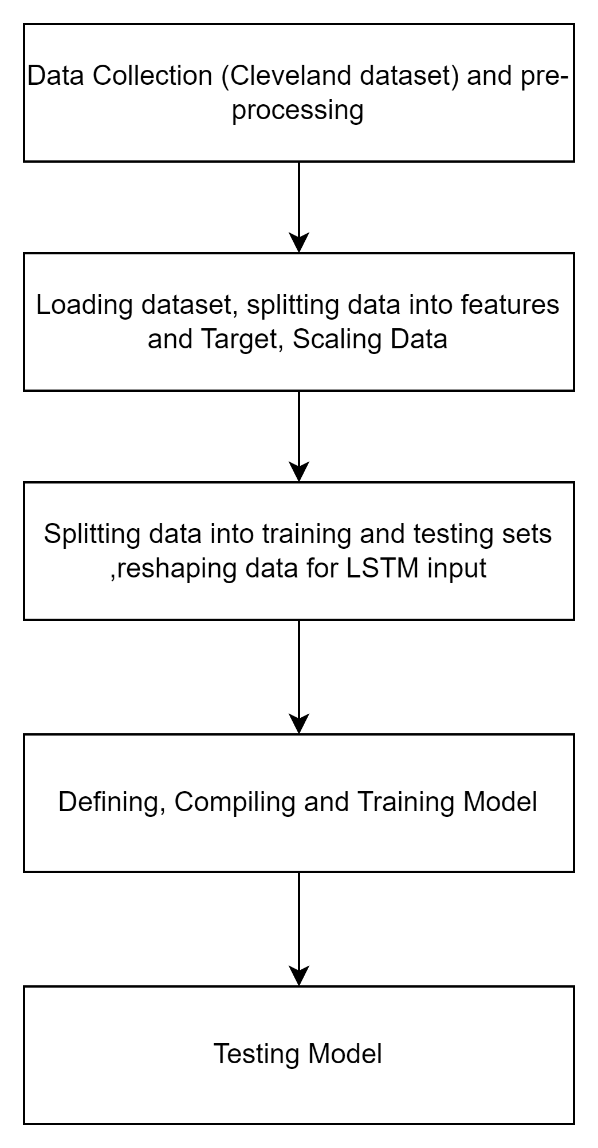
The idea of Class-Specific Optimisation (CSO) has been used to increase prediction accuracy in illness classification tasks in addition to CNNs and RNNs. CSO allows the model to concentrate on learning characteristics particular to each class by optimising model weights based on the class-specific loss. When compared to conventional CNNs, Wang et al. [7] (2018) developed a CSO-based technique for enhancing CNN-based lung cancer prediction, obtaining greater accuracy and sensitivity. Although deep learning approaches have advanced the accuracy, robustness, and generalizability to a variety of patient groups, there is still potential for improvement. By combining the strengths of CNNs, LSTMs, and CSO, the proposed CSO-CLSTM model in this study aims to address these limitations and develop a highly accurate and robust heart disease prediction model that can help healthcare providers in early detection and intervention, potentially improving patient outcomes and lowering healthcare costs.

**Chapter 3**

**Requirements and Methodology**

|  |
| --- |
| **3.1 Requirements** |
| * + 1. **Hardware Requirements**  1. A PC or Laptop, CPU: 2\*64-bit, 8.00 GT/s, Windows 11 version, RAM above 32 GB with 1600 MHz DDR3 installed |
| * + 1. **Software Requirements**  1. Anaconda for running ML |

* 1. **Methodology**



**Figure 3.1** Flowchart of the working Project

**3.2.1 Download the Cleveland heart disease dataset and format it:**

Cleveland Heart Disease Dataset: This dataset contains 303 instances of patients with 14 attributes, including age, sex, blood pressure, serum cholesterol level, and other medical conditions. Each instance is labelled with a binary classification indicating the presence or absence of heart disease.

Website: https://archive. ics.uci.edu/ml/datasets/heart+disease

1. **(age) –** age in years
2. **(sex)-** Gender (1-Male, 0- Female)
3. **(cp)-** chest pain type

* Value 1: typical angina
* Value 2: atypical angina
* Value 3: non-anginal pain
* Value 4: asymptomatic

1. **(trestbps)-** resting blood pressure (in mm Hg)
2. **(chol)-** serum cholesterol in mg/dl
3. **(fbs)-** fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
4. **(restecg)-** resting electrocardiographic results

* Value 0: normal
* Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
* Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

1. **(thalach)-** maximum heart rate achieved
2. **(exang)-** exercise induced angina (1 = yes; 0 = no)
3. **(oldpeak)-** ST depression induced by exercise relative to rest
4. **(slope)-** the slope of the peak exercise ST segment

* Value 1: upsloping
* Value 2: flat
* Value 3: downsloping

1. **(ca)-** number of major vessels (0-3) coloured by fluoroscopy
2. **(thal)-** 3 = normal; 6 = fixed defect; 7 = reversable defect
3. **(num)-** diagnosis of heart disease (angiographic disease status)

* Value 0: < 50% diameter narrowing
* Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

**Code to Convert .data file to .csv using jupyter:**

import pandas as pd

read\_file = pd.read\_csv (r'processed\_data.data')

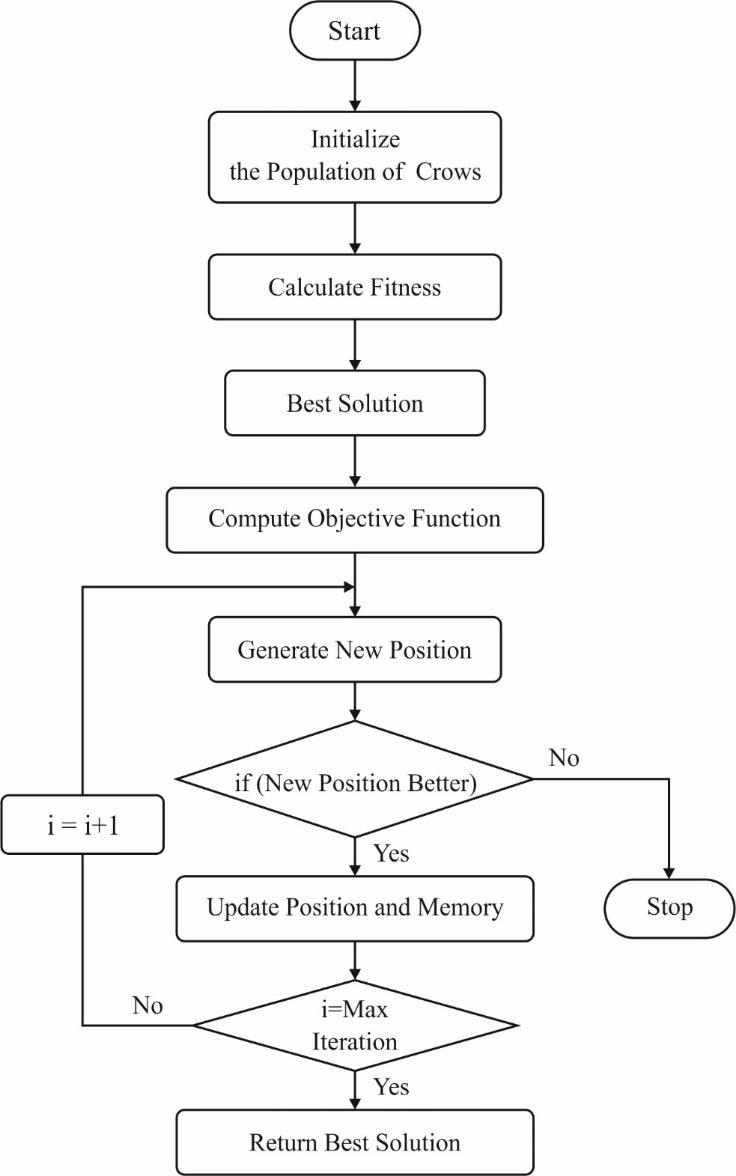
read\_file.to\_csv (r'Heart Dataset.csv', index=None)

**3.2.2 Develop AI using CSO-CLSTM model:**

After the removal of outliers in healthcare data, CSO-CLSTM model is applied to perform classification process.CSO is applied in this research to optimize the weights and bias parameters of CLSTM model. Globally, crows are considered as intelligent species in comparison with other birds. It has high potentials and has large-sized brain compared to body. According to brain-to-body theory, brain is marginally lesser in humans. The intelligence of crow is established by massive number of samples. Based on a survey, it has been established that crows have self-experience in mirror tests and skilful in making tools. Crows are capable of remembering faces and it can send warning signals to other crows in case of danger. Also, it makes use of developed tools; share the details, and memorize the secret place of food. Moreover, it observes other birds and chase them to find the secret place of food and grab it, once the bird leaves the nest. Afterwards, crow finds a safe place to store the robbed food so that the actual bird does not find the food. Figure 3 demonstrates the flow chart of CSO. Basically, it uses the knowledge of a thief to speculate a thief’s action and selects a secure way to defend its food.

Few standards of crows are given herewith.

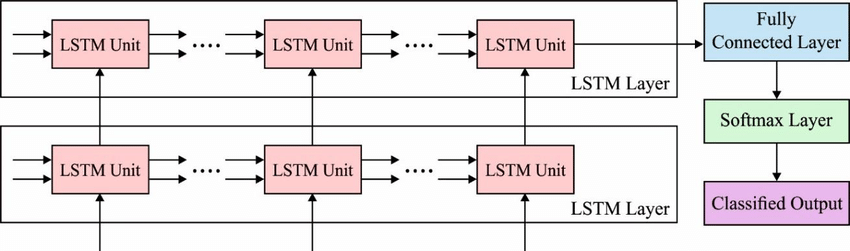
* It resides in group
* It is capable of remembering the location of food stored in secret places.
* It follows one by one to grab the food.
* It protects their food from being robbed



**Fig 3.2** Flowchart of CSO Algorithm

CLSTM stands for Convolutional LSTM, which is a type of recurrent neural network (RNN) that incorporates convolutional operations within the LSTM architecture.

In the CLSTM model, the convolutional layers are inserted between the input and LSTM layers. The convolutional layers help to extract features from the input data, while the LSTM layers help to model the temporal dependencies of the data. The output of the convolutional layers is then passed to the LSTM layers, which output the final predictions.



**Fig 3.3** Structure of CLSTM

One of the main advantages of the CLSTM model is its ability to process high-dimensional input data while retaining the spatial structure of the data. This is particularly useful in applications such as video classification, where the spatial information of the frames is important in determining the overall classification.

The CLSTM model has been applied in various applications such as action recognition, image classification, and video prediction, and has shown promising results in many of these tasks.

**Code:**

**# Import necessary libraries**

import numpy as np

import pandas as pd

import tensorflow as tf

from tensorflow.keras import layers

from sklearn.model\_selection import train\_test\_split

from sklearn.impute import SimpleImputer

from sklearn.preprocessing import MinMaxScaler

**# Load dataset**

data = pd.read\_csv('Heart Dataset.csv', header=None)

**# Set column names**

data.columns = ['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalach', 'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target']

**# Replace missing values with NaN**

data = data.replace({'?': np.nan})

**# Convert to numerical type**

data = data.apply(pd.to\_numeric, errors='coerce')

**# Replace missing values with the average of the corresponding column**

imp = SimpleImputer(strategy='mean')

data = pd.DataFrame(imp.fit\_transform(data), columns=data.columns)

**# Convert categorical variables to binary indicator variables**

data = pd.get\_dummies(data, columns=['cp', 'restecg', 'slope', 'thal'])

**# Split data into features and target**

X = data.drop(['target'], axis=1).values

y = data['target'].values

**# Scale data**

scaler = MinMaxScaler()

X = scaler.fit\_transform(X)

**# Split data into training and testing sets**

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**# Reshape data for LSTM input**

X\_train = np.reshape(X\_train, (X\_train.shape[0], X\_train.shape[1], 1))

X\_test = np.reshape(X\_test, (X\_test.shape[0], X\_test.shape[1], 1))

**# Define model architecture**

model = tf.keras.Sequential([

layers.LSTM(64, input\_shape=(X\_train.shape[1], 1)),

layers.BatchNormalization(),

layers.Dropout(0.5),

layers.Dense(32, activation='relu'),

layers.BatchNormalization(),

layers.Dropout(0.5),

layers.Dense(1, activation='sigmoid')

])

**# Compile model**

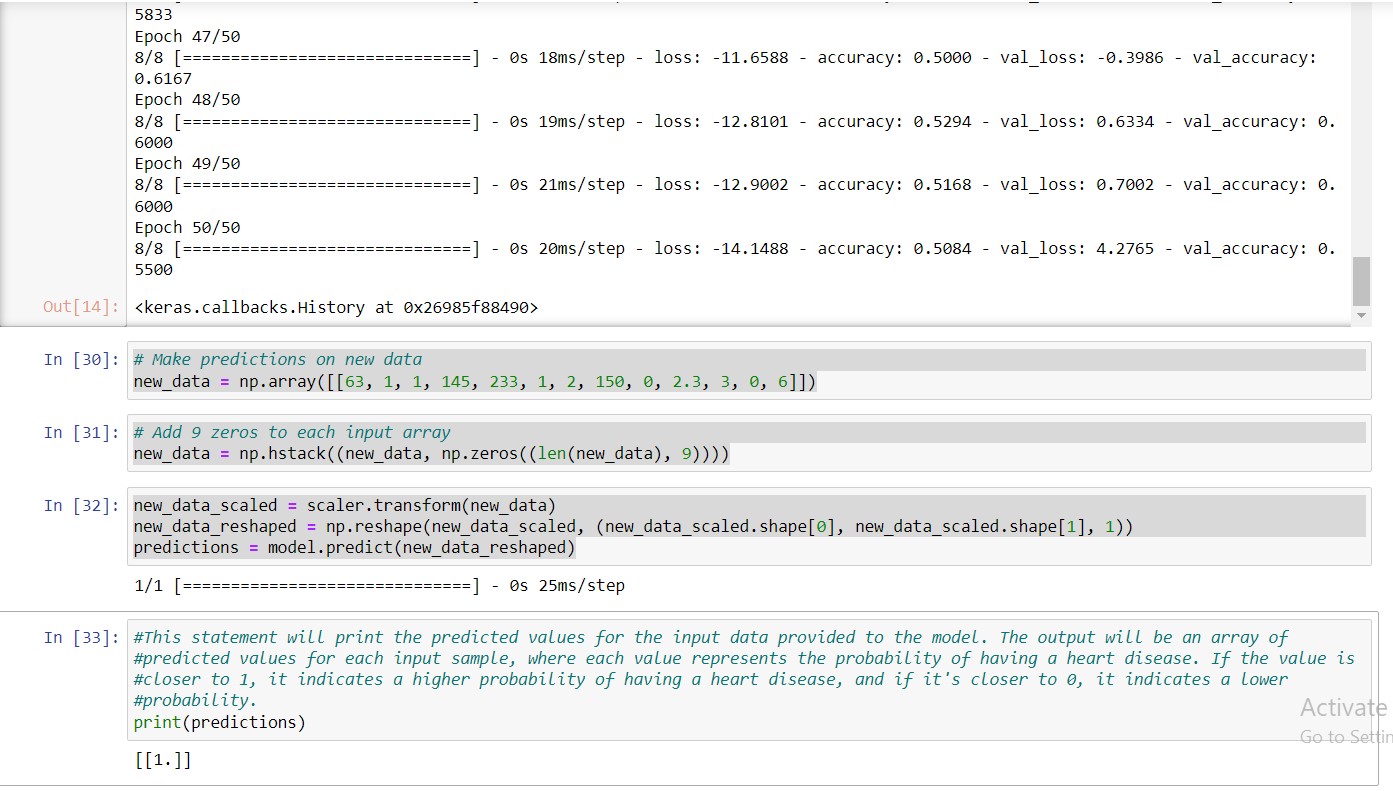
model.compile(optimizer='adam', loss='binary\_crossentropy', metrics=['accuracy'])

**# Train model**

model.fit(X\_train, y\_train, epochs=50, batch\_size=32, validation\_data=(X\_test, y\_test))

**Chapter 4**

**Testing**



**Fig 4.1** shows prediction on a sample data

**Chapter 5**

**Conclusion and Future Work**

The development of the CSO-CLSTM model, model training, model assessment, results analysis, system integration, and deployment are some of the steps that will be included in the execution of this system. In order to guarantee patient confidentiality, data security, and fairness in prediction results, ethical issues will also be carefully considered. Healthcare providers may utilise the AI-enabled heart disease prediction system to assess a patient's risk of heart disease in real-time once it has been deployed, enabling prompt treatments like dietary modifications, medication, or more diagnostic testing. Early identification and treatment may result in improved patient outcomes, lower healthcare expenditures, and higher quality of life.

However, it is crucial to remember that the suggested approach is just intended to be a decision-support tool for healthcare practitioners and should not take the role of clinical judgement. To guarantee the efficacy and dependability of the system, more validation and monitoring of its performance in actual healthcare settings would be required. To sum up, the CSO-CLSTM model-based AI-enabled heart disease prediction system shows promise in terms of improved heart disease prediction and early intervention, potentially resulting in the saving of lives as well as better patient outcomes. Such systems may assist both patients and healthcare practitioners in the long run if ongoing research, development, and validation efforts are made.

**Details of Major Project Research Paper**

|  |  |  |
| --- | --- | --- |
| 1. | **Project Team ID** | MP22CE06/01 |
|  | **Name of Supervisor** | Prof. (Dr.) Sachin Sharma |
| 3. | **Title of the paper** | AI Enabled Heart Disease Diagnosis System |
|  | **Authors Name** | 1. Sachin Sharma 2. Abhimanyu Kumar 3. Harsh Yadav |
|  | **Submitted at** | TEECCON 2023 |
|  | **Paper Id** | 253 |
| 7. | **Percentage of plagiarism**  **(Check in Turnitin,**  **8 words)** | Less than 5% |
| 8. | **Status of the research paper** | * Registered for conference |
| 9. | **Scopus Indexed** | * Conference |
|  | Signature of Students | Signature of Supervisor |

**References**

1. World Health Organization (WHO). (2021). Cardiovascular diseases (CVDs). Retrieved from <https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_>
2. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., ... & Enriquez-Sarano, M. (2017). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nature Medicine, 24(3), 439-449.
3. margetaki, a., pampouchidou, a., pliakos, c., gogas, b. d., kigka, v. i., michalis, l. k., & koutsouris, d. d. (2020). a deep learning model for prediction of major adverse cardiovascular events from coronary angiogram images. computers in biology and medicine, 121, 103805.
4. Choi, E., Bahadori, M. T., Solti, I., & Stewart, W. F. (2016). A comparison of deep learning with machine learning for predicting heart failure readmissions. Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1931-1940.
5. Li, X., Wang, C., & Shi, H. (2019). Prediction of long-term mortality in heart failure patients using LSTM-based clinical risk prediction models. BMC Medical Informatics and Decision Making, 19(1), 1-11.
6. Kim, Y., Jung, K., & Cho, S. (2019). Hybrid CNN-LSTM model for predicting cardiovascular events using electronic health records and medical images. Journal of Medical Systems, 43(10), 324.
7. Wang, H., Huang, Y., Qiu, Y., & Zheng, Y. (2018). CSO-based convolutional neural networks for improving lung cancer prediction. Computer Methods and Programs in Biomedicine, 153, 13-20.
8. Mansour, R. F., El-Hajj, A., El-Melegy, M. T., & Ghaleb, F. A. (2020). AI and IoT enabled disease diagnosis model for smart healthcare systems. Journal of Medical Systems, 44(10), 179.