**Mini Project Report on**  


**DISEASE PREDICTION USING DEEP LEARNING**  


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

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Sohard Behl 2019151**

***Under the Mentorship of***

**Dr. Manoj Diwakar**

Associate Professor, Department of Computer Science

Logo

Description automatically generated

**Department of Computer Science and Engineering**

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

**Dehradun, Uttarakhand**

**July 2023**

A picture containing text

Description automatically generated

**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Disease Prediction using Deep Learning”** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. Manoj Diwakar, Associate Professor** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name University Roll no **signature**

Sohard Behl 2019151

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 4-6 |
| Chapter 2 | Literature Survey | 7-9 |
| Chapter 3 | Methodology | 10-12 |
| Chapter 4 | Result and Discussion | 13-15 |
| Chapter 5 | Conclusion and Future Work | 16-17 |
|  | References | 18 |

**Chapter 1**

**Introduction**

This mini project is an implementation of Deep Learning that is a subset of Machine Learning. It makes use of Neural Networks to make this project work. The primary objective of this project is to develop a robust and accurate predictive model that can effectively identify individuals at risk of heart disease based on various medical and lifestyle factors.

* 1. **Abstract**

Heart disease remains a leading cause of mortality worldwide, emphasizing the need for effective early detection and preventive measures. The rapid advancements in machine learning and artificial intelligence have paved the way for innovative approaches in healthcare, particularly in the field of predictive analytics. In this project report, a comprehensive study on the application of deep learning techniques for heart disease prediction has been presented.

**1.2 Problem Statement**

* Developing an accurate deep learning-based model to predict the risk of heart disease: The primary challenge is to design and train a deep learning model that can effectively analyze complex patterns and relationships within the data to improve prediction accuracy.
* Optimizing deep learning model architecture for determining the best combination of layers to maximize the model's performance.
* Thus, the main purpose of this very project is to predict heart disease with good accuracy and precision.

**1.3 What is Deep Learning?**

* Deep learning is a subfield of machine learning that focuses on training artificial neural networks to learn and make intelligent decisions.
* Deep learning utilizes artificial neural networks with multiple layers to process and extract meaningful representations from input data.
* It evaluates the data iteratively and automatically identifying relevant characteristics and information hierarchy, this enables them to make precise and accurate predictions or classifications.
* These networks are inspired by the structure and function of the human brain, with interconnected nodes or neurons organized in layers.
* The strength of deep learning is its ability to learn and discover detailed patterns from data (raw data) without requiring explicit programming.
* It has numerous applications in many fields such as NLP, computer vision and speech recognition.
* Deep learning models require large amounts of data to effectively learn complex patterns.

**1.4 Objective**

* Heart diseases are the biggest cause of death in the world, claiming around 17.9 million lives each year, with the maximum of the deaths occurring in underdeveloped nations and those with weak healthcare systems.
* The primary objective is to create a predictive model using deep learning techniques that can accurately assess the risk of heart disease.
* The project aims to contribute to the early detection of heart disease by identifying individuals at high risk.
* The model is tested and trained on a dataset, in this project we are using a from the UCI repository and by considering 14 essential properties validating the model.
* The model should be able to determine whether they are having a heart disease accurately and precisely.

According to the World Health Organization (WHO), heart disease is the leading cause of deaths worldwide, taking approximately 17.9 million individuals each year. Heart diseases are a set of heart and blood vascular ailments. More than four out of every five heart disease-related deaths are caused by heart attacks and strokes, and one-third of these deaths occur in individuals under the age of 70. Unhealthy food, physical inactivity, cigarette smoking, and high or harmful alcohol consumption are the most major behavioral risk factors for heart diseases and strokes. Individuals may experience elevated blood pressure (BP), elevated glucose present in blood, elevated lipids present in blood, and overweight or obesity because of behavioral risk factors. These "intermediate risk factors" can be assessed in primary care settings and suggest an elevated risk of heart attack, stroke, heart failure, and other consequences.

Heart diseases are somewhat diverse which can result in a variety of repercussions which includes poor quality of life or even death, particularly in developing nations/countries. Furthermore, the number of people dying from heart disease is higher in developing countries and in countries with worse healthcare systems. This emphasizes the importance of developing a handy system that can guarantee an accurate and early prediction of patients' risk of heart disease.

Tobacco cessation, decreasing the amount of salt in the diet, eating more fruits and vegetables, daily exercise, having balanced diet and avoiding harmful alcohol consumption have all shown to reduce the risk of heart disease. Health policies that encourage conditions that make healthy options affordable and accessible are crucial for motivating individuals to acquire and maintain healthy behaviors.

Thus, the main aim of the project is to enable a user to check whether a person is suffering from a heart disease or not by taking various factors into its consideration such as age, sex, cholesterol level, blood sugar level, heart rate etc. and with these factors the model will make the predictions.

In this project the data used is taken from the UCI dataset with around 1025 rows and 76 properties. From these 76 properties 14 properties are selected which are necessary to validate the model. Potential reasons which can cause heart disease comprises of high blood pressure, diabetes, high cholesterol, obesity, smoking, and a family history of heart disease. Gender, inherited characteristics, and age are the key elements that cannot be changed. Thalassemia is another element in this data set that is determined by genetic factors. Other risk factors include high blood pressure, smoking, high blood cholesterol, a lack of physical activity, being overweight, being unwell, having a high stress level, drinking alcohol, and eating an irregular diet. The various parameters that are used for the deep learning model are age, sex, cholesterol, blood pressure, etc. All these parameters are essential for accurate prediction as it is this data only that is used to train the deep learning model.

**Chapter 2**

**Literature Survey**

In today’s time heart disease is among one of the most severe diseases. An early diagnosis of the heart disease is essential for healthcare expert such that they can easily predict or get to know if a person is at a risk of having a heart disease and can help their patients from contracting the heart disease and saving their lives. For many years, several data analytics methods have been put out to help/assist medical practitioners / medical experts in identifying most heart diseases early signs.

In healthcare groups, biomedical fields, etc., to extract information from the supplied data that is useful. Accurate medical database analysis aids in disease early detection, patient treatment, and social assistance. Machine learning techniques have been effectively applied in a variety of areas, including the prediction of diseases. The goal of creating a classifier system utilizing machine learning algorithms is to significantly aid in the resolution of health-related problems by supporting doctors in early disease prediction and diagnosis.[1]

By examining heart sounds, Luca Brunese in the research paper proposed a system for diagnosing cardiovascular disease. Doctors can now get heart sounds from patients utilising technology and a digital stethoscope, incorporating smart devices like smartphones or tablets that are outfitted with an appropriate mobile application. To determine whether a heart murmur comes from a suffering person with a cardiovascular infection or from a healthy person, the cardiac sounds recorded from patients are converted into numeric values and used as inputs for a deep learning classifier. With an accuracy of 0.98, the suggested method has successfully distinguished between healthy and affected patients.[2]

Anandhavalli Muniasamy et al.'s suggestion that the healthcare industry has challenges in crucial areas such as disease prediction, electronic record administration, computer-aided diagnostics, and data integration. The wellness of any person depends on both lower healthcare costs and a move towards individualised care. Predictive analytics and deep learning, two rapidly expanding technological trends, have begun to play a significant role in the development of massive amounts of medical care information practises and research (Fig. 2.1). Deep learning offers several tools (TensorFlow, Keras), techniques, and frameworks to address the problems in classification, perception, understanding, discovering, and creation. To forecast heart illness, CNN, DNN, and RNN types. Data can be transformed into useful information using these methods. This useful information can be used for training the model and then use it for making predictions whether a person is having heart illness/disease or not.[3]

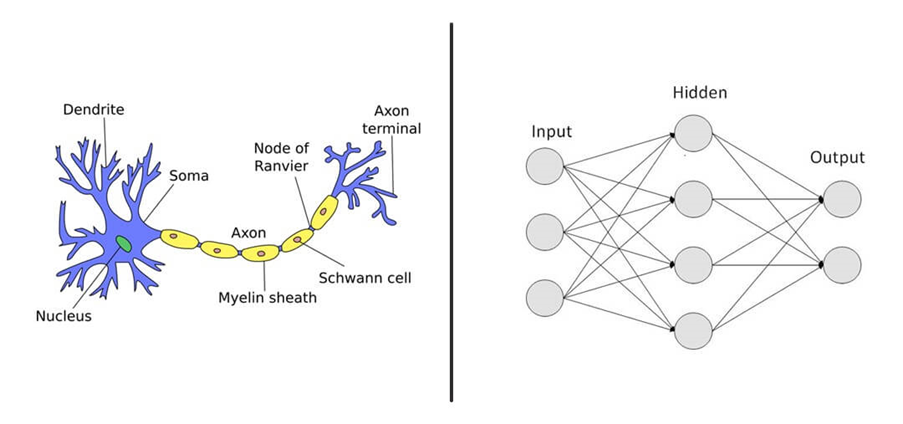


*Fig. 2.1. A representation of Deep Learning with two hidden layers***.**

K-Nearest Neighbour (KNN), Decision Trees, and Support Vector Machines (SVM) were said to be used to evaluate the exactness levels of various machine learning and deep learning methods, according to Syed Nawaz Pasha et al. [4]. The performance and accuracy of the earlier algorithms was reduced when used with large datasets for prediction, so they used TensorFlow and Keras libraries and Artificial Neural Networks (ANN) to increase prediction accuracy.

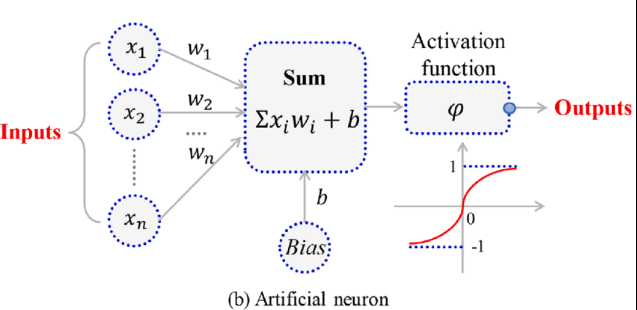
* With above discussions we find that various neural networks are used to predict heart disease or heart illness therefore, figure 2.2 shows us how neurons present in artificial neural networks are like the neurons present in human brain.

A diagram of a neuron

Description automatically generated

*Fig. 2.2. A Representation of Biological Neuron and Artificial Neural Network*

* Now, fig 2.3 shows how these artificial neural networks work when they are provided and input.



*Fig. 2.3. Working of Artificial Neural Network*

We have also taken guidance regarding various latest concepts from discrete sources and knowledge about the proper and efficient usage of different python libraries from the official websites respectively.

**Chapter 3**

**Methodology**

Machine learning is a subset of artificial intelligence where different algorithms enables the system to analyse the data, learn the data patterns and make predictions. Deep learning is specialized form of machine learning i.e., deep learning is a subset of machine learning which uses neural networks with multiple layers to extract complex representations from data for tasks which are very complex/difficult.

Flowchart 3.1 shows various learning techniques and categories of machine learning and deep learning.

***Learning Technique***

Cluster Analysis

Classification

Association Analysis

Regression

*Flowchart 3.1 Learning Techniques and Categories*

The following methodology is provided to develop a heart disease prediction model: -

**3.1 Gathering the requirements.**

This step involves the procedure to figure out all the packages, libraries and the functionality which are needed to build our model properly and accurately. This also requires a proper environment in which the model (code) will run and will give appropriate results.

**3.2 Importing essential Libraries**

Once we figure out all the requirements which are needed, we start building our model on heart disease prediction. Table 3.1 shows all the required libraries we installed.\

|  |  |
| --- | --- |
| ***Library Name*** | ***Installation Command*** |
| NumPy | pip install numpy |
| Panda | pip install panda |
| Matplotlib | pip install matplotlib |
| Scikit-learn | pip install scikit-learn |
| Standard-scaler | pip install standard-scaler |
| TensorFlow | pip install tensorflow |
| Keras | pip install keras |

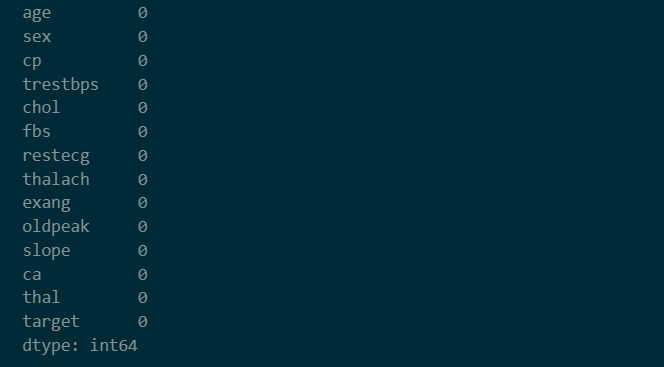
*Table 3.1 Libraries Required*

After installing the above libraries, we import them in our code so that we can work with these libraries and build our model (refer table 3.1).

**3.2 Building the Model**

After importing all the libraries required for the model, we can now start building the model. First of all, we have to read the dataset or import it into our model.

We can check the basic details about the data by running some commands. After that we can check the shape of the dataset, shape refers to the how many rows (instances) and columns (attributes) constitute of the dataset.



*Figure 3.1 Attributes of Dataset*

* Now we can calculate different statistical metrics like mean, standard deviation, mode, median (may be used further). We have to count how many positive and negative targets are there in the all instances and then “target” attribute will be dropped from the dataset.
* The dataset will be divided into 2 datasets namely, “Training set” and “Testing set” in the ratio of 80:20. We can also give any random state to the splitting in order to replicate the same results in another system.
* Now Standard scaler function will be imported from Scikit learning library and also fit & transform the training and testing set by it.
* Keras will be used from tensorflow library to flatten and dense the dataset and it will be assigned to model and this model will be compiled. Furthermore, the neural network model will be trained.
* After all this we can see our result or prediction and the results can also be plotted using matplot library. The plotted graphs show the accuracy and loss of the model compared to the number of epochs i.e., one complete iteration of the training data regarding the prediction of the inputted data.
* The prediction is given in binary i.e., 0 or 1. 0 means no heart disease and 1 means heart disease.
* The model can be further optimized if any kind of modification is required after analysing the result of the neural network model.

**Chapter 4**

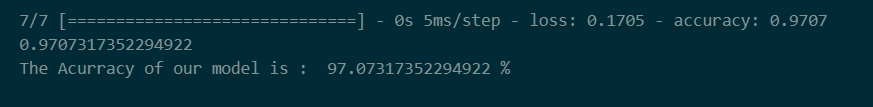
**Result and Discussion**

Now, in this section we will discuss what result has been achieved by the project and what all inferences we can take from this result.

**4.1 Prediction accuracy**

The continuous running of the Deep Learning model shows a range of accuracy. The accuracy of the neural network lies between 93% to 97% hence an average accuracy of 95% which is satisfactory considering the dataset. We can also increase the accuracy by changing the hyperparameters in the model that will change the processing of data by the model.

A higher accuracy can also be achieved by using a greater number of attributes which means having more number of properties or characteristics for the model to work upon. By having a greater number of points, the training of the model will be different and it will be even more accurate in predicting the status of the patient.



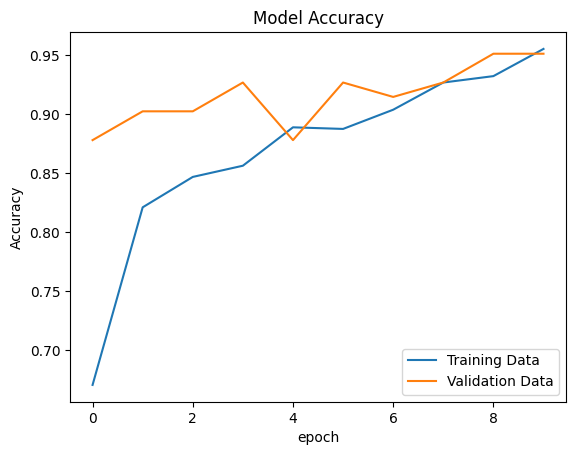
*Figure 4.1 Accuracy of Model*

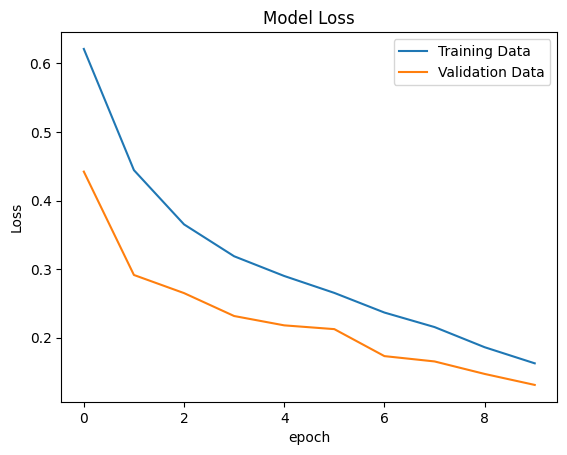
**4.2 Plotting the graphs**

As we have discussed above in the methodology that our model also plots the graphs using matplot library between the number of epochs and accuracy as well as loss of our model.

The first plot shows the relationship between Model accuracy and epoch. The orange line represents the validation data which is also known as testing data and the blue line shows training data, the data on which the model has been trained. The upward trend shows that the accuracy of our model increases as the number of epochs or the number of times our model goes forward or backward in the dataset increases. The validation and training data lines ae close to each other, it means the model is properly fitted on the dataset.

The other plot is of Model loss and epoch. The downward trend shows that the loss decreases as the number of epochs increase.

****

****

**4.2 Inference and Discussion**

As we have discussed in the result section, the accuracy and loss of our model along with the plotting of the graphs. Now we shall discuss about some inferences that can be derived from the achieved result.

**4.2.1 Generalization**

Generalization in the context of deep learning refers to the ability of a trained model to perform well on unseen or new data that it has not encountered during the training phase. A model with good generalization can make accurate predictions on data from the same distribution as the training data, even if the specific samples were not part of the training dataset.

**4.2.2 Overfitting and Underfitting**

Overfitting and underfitting are common issues encountered when training deep learning. Both refer to different types of modelling errors that can impact the generalization ability of a model.

Overfitting means that the learning of training data by the model is too well but it does not perform on test data or validation data. The models can not generalize to new data.

It happens when the model is too complex, has many parameters and trained for too many epochs.

Both overfitting and underfitting are not good for our model.

Overfitting can be removed by stopping early i.e., monitoring the model's performance on a validation set and stopping training when the performance starts to degrade. Augmenting the data can also help. Implementing regularization is used very often for the same.

Underfitting can be dealt by increasing the complexity of the model or by tuning the hyperparameters.

When we strike the right balance between avoiding underfitting and overfitting, our model has an increased accuracy and it can be used efficiently.

Cross-validation techniques like k-fold cross-validation are used for improving generalization in the deep learning model.

Now in the next chapter, we shall conclude this project report and also discuss future works that can be done regarding this disease prediction deep learning project.

**Chapter 5**

**Conclusion and Future Work**

**5.1 Conclusion**

In conclusion, the application of deep learning for heart disease prediction shows promising results in enhancing early detection and intervention strategies. By leveraging the power of artificial intelligence and machine learning, we can extract meaningful patterns and features from extensive datasets, enabling accurate predictions and personalized healthcare management. This project report highlights the importance of comprehensive data collection, pre-processing, and model optimization to ensure robust and reliable predictions.

By building this deep learning project, we have learnt about deep learning techniques and different categories like cluster analysis, association analysis, classification, regression  
  
The successful implementation of this deep learning-based heart disease prediction system holds significant potential for improving patient outcomes and public health. By identifying individuals at high risk of developing heart disease, healthcare professionals can intervene early, provide timely treatment, and implement targeted preventive measures.

We can conclude that our project is highly accurate and can be developed further for more improvements and even more accurate predictions.

This project has a lot of potential in the upcoming future and will be developed further. Future works that are possible in the deep learning based-heart disease prediction system are discussed in the upcoming section.

**5.2 Future Work**

There are numerous possibilities in future where this deep learning project can be implemented and we can work on further development of this project to imply it on other diseases also by training the model on the respective datasets.  
Although this project has made tremendous progress in utilizing deep learning to predict heart disease, there are still a number of areas that are worth more study and development:

* Expanding the dataset used for training and validation can enhance the generalizability and robustness of the predictive model. Incorporating data from multiple sources and populations can provide a more comprehensive understanding of heart disease risk factors.
* Exploring advanced techniques for feature selection and extraction can help identify the most relevant and informative variables for heart disease prediction. This can potentially improve model performance and reduce computational complexity.
* Integrating deep learning-based prediction models into health record systems can facilitate seamless implementation and utilization by healthcare professionals. This can enhance the real-world impact of the deep learning model and enable their practical application in healthcare settings.
* Developing methods for interpreting and explaining the predictions of deep learning models can enhance transparency and trust in the system.
* Developing the model in context of IoT for supporting day to day life of people such that it becomes easily accessible to everyone through electronic devices as well as mobile applications, image processing can also be introduced into this deep learning model.

By addressing these areas of future work, researchers and healthcare practitioners can continue to advance the field of heart disease prediction using deep learning, ultimately leading to improved patient care, preventive strategies, and better management of cardiovascular health.

**References**

[1] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130.

[2] Luca Brunese, Fabio Mart inelli, Francesco Mercaldo, and Antonella Santone. 2020. Deep learning for heart disease detect ion through cardiac sounds.

[3] Anandhavalli Muniasamy, Sehrish Tabassam, Mohammad A. Hussain, Habeeba Sultana, Vasanthi Muniasamy, and Roheet Bhatnagar. 2020. Deep Learning for Predictive Analytics in Healthcare.

[4]Syed Nawaz Pasha, Dadi Ramesh, Sallauddin Mohmmad, A. Harshavardhan, and Shabana. 2020. cardiovascular disease prediction using deep learning techniques.

[5] Ramesh TR, Umesh Kumar Lilhore, Poongodi M, Sarita Simaiya, Amandeep Kaur, Mounir Hamdi. (2022). Predictive Analysis of Heart Diseases with Machine Learning Approaches.

[6] About Neural Networks <https://en.wikipedia.org/wiki/Neural_network>

[7] Python libraries: <https://pypi.org/>

[8] Deep Learning models: <https://towardsdatascience.com/6-deep-learning-models-10d20afec175>