Experiment No. 1
Review of Deep Learning techniques
Date of Performance:
Date of Submission:

Experiment No. 01

Paper – 01: Heart disease risk prediction using deep learning techniques with feature augmentation

Problem Statement:

The problem addressed in this study is the early detection of heart problems, which are a significant risk to public health. The primary aim of this research is to advance the field of heart disease classification by attaining exceptionally high success rates in early disease detection. In this pursuit, the study also sets out to address two distinct yet interrelated secondary goals. Firstly, it seeks to establish a novel methodology capable of effectively addressing classification challenges, particularly those associated with datasets featuring a limited number of features. Secondly, the paper explores the potential of convolutional neural networks to refine and augment existing techniques used for feature engineering and augmentation. By addressing these objectives, the research aims to significantly enhance the accuracy and efficiency of heart disease classification, ultimately contributing to improved healthcare outcomes and the well-being of a large population.

Solution:

The solution proposed in this paper combines convolutional neural networks (CNNs) and sparse autoencoders to improve the early detection of heart diseases. This approach involves data preprocessing with CNNs, allowing the extraction of relevant features. The sparse autoencoder refines these features, enhancing their informativeness and reducing noise. The integrated architecture classifies patients based on these processed features, achieving a remarkable precision of up to 90%. This precision level represents a significant advancement in cardiovascular disease risk assessment. The innovative methodology also addresses classification problems with minimal features and harnesses the power of CNNs for feature augmentation. This research contributes to the improvement of healthcare outcomes, potentially saving lives among a large population at risk of heart diseases.

Technologies:

1. **Sparse Autoencoder (SAE):** A neural network architecture used for feature augmentation. SAEs are trained to learn sparse representations of the input data, reducing dimensionality and enhancing the informativeness of features.
2. **Machine Learning Classifiers:** The paper evaluates two types of classifiers, which are connected to the SAE for the classification task:
    - **Multilayer Perceptron (MLP):** A traditional feedforward neural network used for classification tasks.
    - **Convolutional Neural Network (CNN):** A deep learning architecture typically used for image-related tasks, including feature extraction and classification.

3. **Feature Augmentation:** The paper explores various approaches for dealing with data characterized by a small number of features. It includes techniques for increasing the number of features, and this augmented dataset is used for classification.
4. **Parallel Training:** The proposed approach involves a parallel training process, where a data augmentation neural network (Sparse Autoencoder) and a classifier network are trained simultaneously. The augmented dataset is used as input to the classifier network to predict the final class.
5. **Deep Learning:** The use of deep learning techniques is evident in the incorporation of convolutional neural networks, which are particularly well-suited for extracting hierarchical features from data.
6. **Data Processing:** The paper emphasizes the importance of processing and improving information to achieve more accurate predictions, indicating a focus on data processing techniques

Dataset:

The dataset used is made up of 11 clinical features: the patient's age, sex, type of chest pain (typical angina, atypical angina, non-anginal pain or asymptomatic), the resting blood pressure mmHg, the serum cholesterol (mm/dl), the fasting blood sugar (value 1 if FastingBS > 120 mg/dl, and value 0 otherwise), resting electrocardiogram results (which can be Normal, ST if the patient has ST-T abnormalities or LVH if the patient shows probable ventricular hypertrophy), the maximum heart rate (numeric value between 60 and 202), exercise-induced angina which can be yes or no, the oldpeak (numeric value measured in depression) and finally, the slope of the peak exercise ST segment (Up, Flat, Down). The column number 12 contains the output class which can be 1 (heart disease) or 0 (normal).

Conclusion:

The paper introduces an innovative deep learning approach for predicting heart problems in a dataset with limited features. It leverages both Sparse Autoencoder and Convolutional Classifier to enhance feature extraction and classification. Through joint training, the Convolutional Neural Network surpasses the Multilayer Perceptron by 0.6% in accuracy. Optimal feature space size is determined as 200 neurons. This method offers a substantial 4.4% improvement over traditional classifiers and outperforms complex stacking techniques found in the state-of-the-art. Given the life-saving implications of early heart problem detection, these advancements hold great promise for healthcare specialists.

Paper – 02: Heart attack prediction using deep learning techniques

Problem Statement:

The problem statement of this research paper is to develop an effective deep learning model, specifically utilizing Convolutional Neural Networks (CNNs), for the early prediction of heart attacks. Heart attacks, also known as myocardial infarctions, are a leading cause of death globally and often occur suddenly without warning. Timely prediction is vital for early intervention and prevention. This project aims to harness the power of CNNs in medical imaging and data analysis to predict the risk of a heart attack based on relevant diagnostic images and patient data.

Solution:

The paper proposes a system, with a strong prediction algorithm, which implements powerful classification steps with a comprehensive report generation module. The project aims to implement a self learning protocol such that the past inputs of the disease outcomes determine the future possibilities of the heart disease to a particular user. The proposed model makes use of strong preprocessing tools so that the classification and prediction do not show any errors relating to the dataset. This paper aims to address these and propose implementation of innovative features to develop a more comprehensive system.

Technologies:

This deep learning model has a simple structure with a built-in feedback loop allowing it to act as a forecasting engine. RNN in essence is a regular neural network with an additional hidden state where the hidden state influences the neural network output. The hidden state is updated on each input step. It is a model which can not only learn local and temporal dependencies in data but also can accommodate variable sequence lengths. There are several ways to address this problem the most popular of which is gating. The most popular of gating types are LSTM and GRU with GRU being used in the paper.

Dataset:

Dataset is 303 records with 75 medical attributes (factors) from the UCI Machine Learning Data Repository [3]. After preprocessing use of 270 records with 13 medical attributes. Out of these 13 attributes, 7 have discrete values whereas 6 have continuous values. The preprocessed data is given as input at the input layer. The word embedding layer converts the input data into dense vector representation. With the help of learning GRU layer the vector values are processed. The softmax activation function present at fully connected layer helps in classifying into heart disease or non-heart disease patient.

Conclusion:

As we have been through number of projects and their papers and found different algorithms had different accuracy starting from ML algorithms to deep learning algorithms accuracy kept on increasing but we were not able to obtain good results for silent heart attack prediction. Hence after analysis we thought of using RNN and GRU to make the system more accurate and efficient to predict the silent heart attacks and inform the user at the earliest possible. This system has increased the heart attack prediction accuracy to 92% and has proved to be an excellent source in predicting silent heart attacks.

Paper – 03: Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators

Problem Statement:

Cardiovascular diseases (CVDs) are a leading cause of mortality worldwide, accounting for a significant percentage of global deaths each year. Timely and accurate prediction of heart disease risk is crucial for effective preventive interventions and personalized healthcare. The problem addressed in this research paper revolves around the development of an innovative and comprehensive heart disease prediction model that incorporates self-measurable physical condition indicators. The challenge lies in effectively integrating heterogeneous data from wearable devices and traditional clinical sources, ensuring data quality, and devising machine learning algorithms capable of extracting meaningful patterns from this diverse dataset.

Solution:

The solution involves a multidisciplinary approach combining aspects of data collection, feature engineering, machine learning algorithms, and validation techniques. With the adoption of Random Forest, the best accuracy of 82.18% has been achieved by modification of feature selection. By following these steps and integrating self-measurable physical condition indicators with advanced machine learning techniques, it is possible to create an effective heart disease prediction model that can revolutionize personalized healthcare and contribute significantly to preventive interventions. It is recommendable to use sophisticated equipment to detect potential heart risks in advance.

Technologies:

1.Data Processing - For data description, the research utilized the describe function and pandas profiling in Python to summarize the dataset.
2. Machine Learning Algorithms –

- Logistics Regression - The Logistic Regression from the sklearn package in Python was used to build the model.
- K-Nearest Neighbors - K-Nearest Neighbors (KNN) is a classification algorithm. The research chose 1 to 20 as the number of neighbors. The K Neighbors Classifier Scores were calculated for each number of neighbors.
- Support Vector Machine - Support Vector Machine was chosen as one of the models because it is an algorithm for classification and regression. The research used svm from sklearn.svm package in Python.
- Decision Tree - Decision tree was chosen because it is a nonparametric machine learning model for classification and regression.
- Random Forest - Random Forest is an algorithm consisting of decision trees. Random Forest Classifier from the sklearn. ensumble package was used to build the home and all matrices models.

Dataset :

Dataset used in this paper is from the Cleveland heart data set from the UCI machine learning repository. The data we selected is made up of 14 variables and 303 instances. Overall speaking, there are 13 variables and 1 categorical response variables (target). Among these variables, numerical variables are age, trtbps, chol, thalach, old peak; Categorical variables are sex, exang, cp, fbs, rest_ecg, slp, thall, target. The table below illuminates the meaning of each variable.

Conclusion:

The paper concludes that the machine learning algorithms with only self-measurable physical condition indicators do not predict as accurately as machine learning algorithms with all physical condition indicators. Not only do algorithms with self-measurable physical condition indicators not predict the heart disease outcome as accurately as algorithms with all physical condition indicators, but they are also more likely to falsely predict not having heart disease among patients with heart disease. Thus, machine learning algorithms with only self-measurable physical condition indicators should not be used until more indicators are measurable at home in the future.

Analysis Table:

Paper	Paper1	Paper 2	Paper 3
Advantage	-It utilizes a Sparse Autoencoder for feature augmentation, enhancing data representation.	-This system has increased the heart attack prediction accuracy and has proved to be an excellent source in predicting silent heart attacks.	- It is possible to create an effective heart disease prediction model that can revolutionize personalized healthcare and contribute significantly to preventive interventions.
Disadvantage	-Increased complexity of the architecture, which can lead to challenges in implementation and hyperparameter tuning.	-Does not provide detailed evaluation of a proposed approach.	- Algorithms with self-measurable physical condition indicators not predict the heart disease outcome as accurately as algorithms with all physical condition indicators, but they are also more likely to falsely predict not having heart disease among patients with heart disease.
Performance	-Contingent on the quality and quantity of the training data, the effectiveness of hyperparameter tuning, and the choice of classifier, making it variable in practice.	-Achieves an accuracy of 92% in predicting silent heart attacks.	-With the adoption of Random Forest, the best accuracy of 82.18% has been achieved by modification of feature selection.
Complexity	-Relatively high due to the combination of sparse autoencoder-based feature augmentation	- Relatively low model complexity	-Moderate complexity
Dataset	-11 clinical features, including patient demographics, chest pain type, blood pressure, cholesterol, fasting blood sugar, ECG results, heart rate, exercise-induced angina, and oldpeak values.	303 records with 75 medical attributes (factors) from the UCI Machine Learning Data Repository. After preprocessing use of 270 records with 13 medical attributes. Out of these 13 attributes, 7 have discrete values whereas 6 have continuous values	-The data is made up of 14 variables and 303 instances. Overall speaking, there are 13 variables and 1 categorical response variables (target).