Brain Stroke Prediction Using Machine Learning Approach

ABSTRACT

Stroke is a medical condition in which poor blood flow to the brain results in cell death. Nowadays stroke has been found by inspecting the affected individuals. Using these risk factors, a number of works have been carried out for predicting and classifying stroke diseases. Most of the models are based on data mining and machine learning algorithms.

In many nations, stroke is the main cause of endomorphy and death. The goal of this research was to figure out how to make things better. We have used the stroke disease data set from Kaggle. Patients can benefit from data that has been pre-processed. Ischemic stroke and stroke hemorrhage are two forms of stroke, Individuals are divided into two categories using machine learning methods.

Machine learning techniques were employed seven times in this investigation. Logistic Regression, Support Vector Machine (SVM), Random Forest, Cat Boost, Naive Bayes, KNearest Neighbors Because of this, our findings, Cat Boost makes the best accuracy, along with precision and recall values, and the f1-Score.

Key Words

Stroke, Machine Learning, Endomorphy, Medical Condition, Prediction, Data Mining.

INTRODUCTION

Atrial fibrillation, a cardiac condition characterized by irregular heartbeats, stands as a prominent precursor to stroke, which can have dire consequences, including fatality. For medical professionals, the arduous and time-consuming task of predicting and preventing strokes is an ongoing challenge. Stroke, a disease that predominantly affects individuals aged 65 and above, inflicts significant damage to the brain, resembling the destructive impact of a "coronary episode." It ranks as the third leading cause of death in the United States and other agricultural nations. The occurrence of a stroke is intimately tied to the impairment of cerebral blood flow, manifesting in two primary categories: ischemic strokes and hemorrhagic strokes. The former results from insufficient blood supply, while the latter is characterized by bleeding within the brain. Both types necessitate prompt and distinct treatments to mitigate their devastating effects.

LITERATURE SURVEY

1. Analyzing the Performance of Machine Learning Algorithms for Stroke Detection (2023)
The application of machine learning algorithms has demonstrated promise in the early diagnosis of strokes in the rapidly changing field of healthcare. In order to gain a better understanding of these algorithms' strengths and limits in recognizing stroke cases, this study examines how well they function.

2. Stroke Prediction Using Machine Learning Algorithms (2023)
The use of machine learning in stroke prediction is becoming more and more relevant as the discipline develops. The viability of using machine learning models to anticipate strokes and facilitate early intervention is investigated in this research study.

3. Brain Stroke Prediction Using Machine Learning Approach (2021)
Machine learning-based brain stroke prediction made significant progress in 2021. This work explores the use of Support Vector Machines, a type of machine learning, to forecast brain strokes. Although support vector machines are employed, it is mentioned that this method might be less precise.

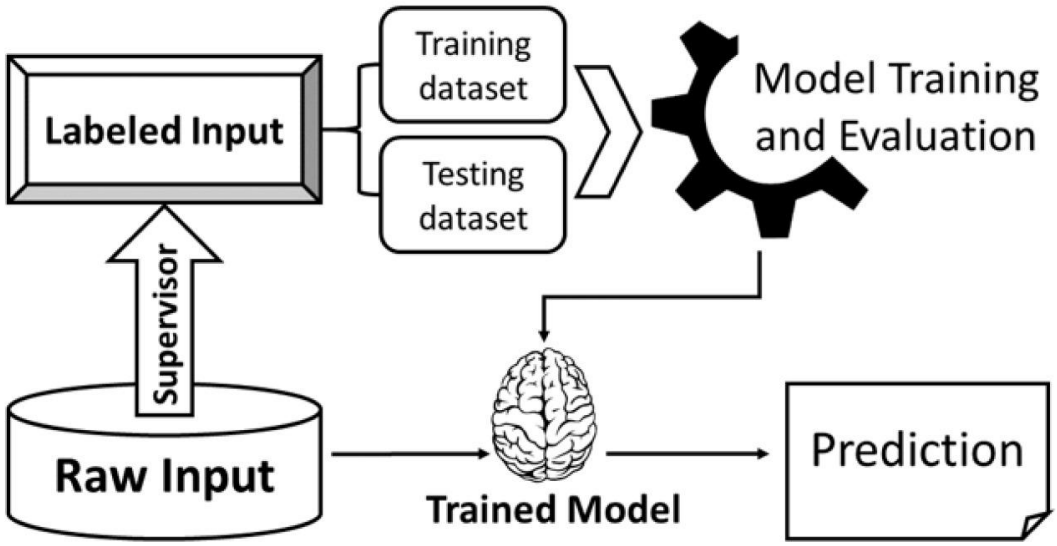
4. Evaluating Machine Learning Algorithms for Stroke Detection (2023)
The use of machine learning techniques has expanded the range of stroke detection options. The performance evaluation of several algorithms, such as K-Nearest Neighbors, Random Forest, and Decision Trees, is the main objective of this study. Some of these algorithms have been found to have false negative rates and lengthy training times.

5. Logistic Regression and Naive Bayes Algorithms for Stroke Prediction (2022)
This study investigates the use of Naive Bayes and logistic regression methods for stroke prediction. The study does point out that these models could be time-consuming and demand additional computer space.

PROPOSED SYSTEM

Here in the proposed system, For finding stroke disease classification, machine learning techniques are K-Nearest Neighbors(KNN),Support Vector Machine (SVM), & other machine learning algorithms like Random Forest, Cat Boost, Logistic Regression, Naive Bayes, and Multi-layer Perceptron.
Cat Boost has got the best accuracy using nowadays that can improve the accuracy to find out the stroke disease.

ARCHITECTURE



Algorithms

1.Random Forest:

- Type: Ensemble Learning
- Features: Merges several decision trees to produce predictions with higher accuracy. It's like combining the wisdom of a crowd to make reliable and effective predictions for both regression and classification tasks.

2.K-Nearest Neighbor (KNN):

- Type: Instance-Based Learning
- Features: Classifies data points based on the majority class of their K-nearest neighbors. Imagine it as connecting data points to their closest neighbors to make decisions, which is particularly efficient with smaller datasets.

3.Naive Bayes:

- Type: Probabilistic Classification
- Features: It assumes features are independent, much like pieces of a puzzle fitting together. This approach works effectively in tasks such as spam and text classification.

4.Support Vector Machine (SVM):

- Type: Discriminative

- Features: Identifies the best hyperplane to separate data into classes, like finding the optimal line to divide points on a graph. SVM is efficient, especially for high-dimensional data in both regression and classification tasks.

5. CatBoost:
- Type: Gradient Boosting
- Features: CatBoost is a specialized algorithm for handling categorical features. Think of it as a well-balanced scale, requiring minimal adjustment, and delivering robust predictive performance.

6.Decision Tree:
- Type: Non-parametric
- Features: Decision trees construct a tree-like structure, making decisions at each branching point, similar to how you make decisions based on choices. They're easy to understand but require careful pruning to avoid overfitting.

7. Multi-Layer Perceptrons (MLP):
- Type: Neural Network
- Features: MLPs are like interconnected layers of neurons in the brain, used for complex tasks like recognizing patterns in images and speech.

IMPLEMENTATION

Step 1: Data Ingestion
The first order of business was to open the dataset by loading it into a suitable data processing environment. The data was imported from the CSV file, ready for further analysis.

Step 2: Data Splitting
With the dataset at hand, the data was strategically divided into two distinct subsets: the training set and the test set. This partition is crucial to evaluate the performance of machine learning algorithms accurately.

Step 3: Algorithm Implementation
The heart of the operation lay in the application of machine learning algorithms to the segmented data. Multiple algorithms were put to work, each aiming to detect and predict bleeding occurrences based on the available data. During this process, algorithmic performances were scrutinized.

Step 4: Cat Boost Excellence
Among the array of algorithms at play, one shone exceptionally bright - Cat Boost. This algorithm displayed the highest accuracy score, signifying its efficacy in bleeding diagnosis. Its robust capabilities in handling categorical features likely played a pivotal role in achieving this feat.

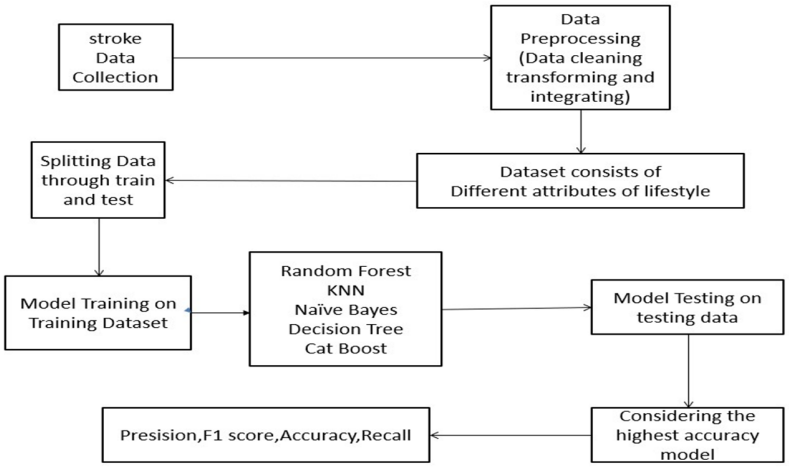
Step 5: User-Friendly Interface

A user-friendly interface was crafted, offering an array of functionalities to the user. These included the ability to explore vital sections like the home page, an informative 'about' section, and the option to access the data split, model performance metrics, predictions, and visual representations in the form of graphs.

Step 6: Flask Framework Integration
To bring the entire operation together seamlessly, the Flask framework was adopted. Flask not only facilitated the development of a web-based interface but also ensured that all these components interact harmoniously. Users could navigate the application effortlessly and, in the end, receive precise bleeding diagnosis predictions.

Data flow diagram

We are utilizing a variety of machine learning methods in this suggested system, including support vector machines, random forests, K-nearest neighbors, logistic regression,cat boost, and decision trees. We compare these several algorithms with one another in our suggested system to choose the model with the highest accuracy. According to the accuracy rating, we will decide which model works best with our dataset.



Result

Conclusion:

In this, we use machine learning methods to classify the stroke data. Then, eight algorithms—Logistic Regression, K-Nearest Neighbors, Naive Bayes, Decision Tree, Random Forest, Multi-layer Perceptron, Deep Learning and Support Vector Machine, and Cat Boost—are compared to perform classification. Our experiment indicates that, when compared to other assessed classification methods, the Cat Boost approach was employed as a classification methodology. The cat surge method provides the highest

degree of accuracy. The precision of the categorization algorithm using theThe value of the default optimization parameter has not been examined nonetheless. The classification model possesses the possibility for improvement right now. To raise the machine learning algorithm's accuracy, parameters need to be adjusted. More individuals than individuals with ischemic strokes and brain hemorrhages.

References:

- Concept of Stroke by Healthline.
- Dataset named ‘Stroke Prediction Dataset’ from Kaggle: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset/code
- Stroke Prediction Using Machine Learning Algorithms: https://ijirem.org/DOC/2-strokeprediction-using-machine-learningalgorithms.pdf.
- G. R. Kumar, P. Vyshnavi, S. Prasanna, T. H. Reddy, C. Charanya, and P. Chandrababu, “Brain Stroke Detection Using Machine Learning", IJREAM, vol. 5, no. 3, pp. 34–36, Mar. 2022.
- Mr.M.THIRUNAVUKKARASU, et. al. International Journal of Engineering Research and Applications
  www.ijera.com ISSN: 2248-9622, Vol. 13, Issue 4, April 2023, pp. 237-242.