MACHINE LEARNING ALGORITHM FOR BRAIN STROKE DETECTION

Ashish Ku. Lakra
Department of Computer Science and Engineering
Chandigarh University
ashishlakra2k2@outlook.com

Monika
Department of Computer Science and Engineering
Chandigarh University
monika.e9911@cumail.in

Md. Raghib
Department of Computer Science and Engineering
Chandigarh University
Mdraghib612@outlook.com

Sidharth Samantaray
Department of Computer Science and Engineering
Chandigarh University
sidharthSamantaray04@gmail.com

***Abstract*–This study presents an innovative hybrid algorithm that integrates the convolutional neural network (CNN) architecture with diverse machine learning techniques to achieve binary classification of authentic CT images depicting brain strokes. Here we have used various feature selection method and machine learning algorithms such as Support Vector Machines, Random Forest, k-Nearest Neighbors, Naïve Bayes, Convolutional Neural Network and k-Nearest Network, significantly enhances accuracy and efficiency in detecting strokes from CT images. The OzNet-mRMR-NB hybrid algorithm achieves outstanding results, with an accuracy of 98.42% and an AUC of 0.99, offering the potential to revolutionize early stroke detection and treatment.**

As stroke stands as a major global cause of death and disability, the development of this machine learning algorithm holds immense promise. It surpasses current methods in terms of accuracy, speed, and portability, making it particularly valuable for remote areas lacking specialized medical equipment. In conclusion, this research shows the transformative potential of machine learning algorithms in advancing the early detection and treatment of brain strokes, ultimately leading to improved patient outcomes.

***Keywords*– CNN, Random Forest, Naive Bayes, kNN, SVM, Logistic Regression**

I. INTRODUCTION

In this research, we address a critical issue concerning the limitations of existing diagnostic methods for the detection of brain strokes [7,17]. The conventional methods often rely on subjective visual interpretation of medical images, a process prone to errors and subjectivity. Additionally, in regions where expert radiologists are in short supply, delays in the diagnosis of brain strokes can have severe consequences. This challenge extends to the pressing need for automated, data-driven solutions that can not only expedite the detection of brain strokes but also improve accuracy. Machine learning algorithms [27] have emerged as a promising avenue to address these issues, capitalizing on extensive datasets to learn intricate patterns [6,8] and deliver consistent and reliable diagnostic results. To tackle this issue effectively, it is vital to develop and validate machine learning algorithms that not only accurately detect brain strokes [10,15] but also offer transparency and interpretability in their decision-making processes. The successful resolution of this issue holds the potential to significantly enhance patient outcomes by ensuring timely intervention and treatment for those at risk of or affected by brain strokes.

The core problem at the heart of our research is the accurate and timely identification of brain strokes in patients. Brain strokes [14] are acute medical emergencies, demanding immediate attention, and any delays in their diagnosis can lead to severe consequences, including disability and, in some cases, death. This problem encompasses several key challenges, beginning with the paramount importance of early detection, as brain strokes can often develop subtly and progress rapidly. Additionally, the intricate patterns [6,8] within brain imaging data, such as MRI [2,4,16] and CT [9,24] scans, pose challenges beyond human observation. Patient data exhibits significant variability, including differences in age, sex, medical history, and risk factors, which the algorithm must account for to provide accurate diagnoses. The accuracy of the algorithm is of utmost importance to minimize false positives and false negatives, which could result in incorrect diagnoses and treatments. To address these multifaceted challenges, our research encompasses a series of essential tasks, from data collection and preprocessing to model development. Machine learning algorithms [27], including Naive Bayes, Logistic Regression, k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest and Convolutional Neural Networks (CNN) [11,12], are harnessed to learn and detect intricate patterns [6,8] within brain images. Rigorous performance evaluation and the interpretation of the algorithm's decisions, ensuring it garners trust and understanding from medical practitioners, are integral aspects of our research.

The research, by harnessing the transformative potential of machine learning and various algorithms, presents a promising solution to revolutionize the early detection and treatment of brain strokes [15]. The synergy of machine learning with advanced imaging techniques, such as the OzNet [1] convolutional neural network [12] architecture, demonstrates exceptional results. Achieving an accuracy of 98.42% and an AUC of 0.99, this algorithm sets a new standard in the realm of brain stroke detection. This advancement is a promising step towards ensuring that brain strokes [15,20] are detected in a timely and accurate manner, thus mitigating their potentially severe consequences. The research underscores the critical role of machine learning in improving patient outcomes by providing rapid, accurate, and consistent diagnostic results, which can be instrumental in ensuring timely intervention and treatment for individuals at risk of or affected by brain strokes.

II. LITERATURE REVIEW

A diverse array of significant studies has significantly contributed to the field of stroke detection [10,20,13] and medical image analysis. For instance, Ozaltin et al. (2022) introduced "OzNet,"[1] a deep learning [3.5] approach customized for detecting strokes in brain CT [9] images, effectively showcasing recent advancements in applying deep learning techniques [24] to this vital medical task. Another noteworthy work by El-Dahshan E.-S.A. et al. (2010) delves into the realm of hybrid intelligent techniques, with a focus on classifying MRI [2,4] brain images. While not tailored exclusively to stroke detection, this research makes a valuable contribution to the broader domain of medical image analysis.

In a similar vein, Al-Tuwaijari J.M. et al. (2020) explores the application of deep learning techniques [5,24] in the context of advancing plant leaf diseases detection. This study underscores the remarkable versatility of deep learning [24] across diverse domains, even when not directly linked to brain stroke [19] detection. Further enriching the landscape of medical image analysis, Saritha M. et al. (2013) center their research on the classification of MRI [2,16] brain images. Employing techniques such as wavelet entropy and probabilistic neural networks, their work introduces a relevant method for brain image analysis.

Li L. et al. (2020) focus on the realm of deep learning, specifically in the context of detecting and segmenting hemorrhagic lesions in brain CT images [3,5,24]. Their research highlights the practical application of deep learning in addressing specific aspects of stroke detection. In the domain of deep learning and convolutional neural networks (CNNs) [11,12], Szegedy C. et al. (2015) have made significant contributions by introducing the foundational concept of "Going deeper with convolutions." This pioneering work plays a crucial role in advancing deep learning models for medical image analysis [24].

Zhu G. et al. (2020) introduce a novel method for classifying strokes, especially in simulated electromagnetic imaging, using graph approaches. Their study demonstrates the varied techniques used in stroke detection. In a separate study, Kasabov N. et al. (2014) investigate evolving spiking neural networks, emphasizing personalized modeling, classification, and prediction of spatio-temporal patterns [6,8], including strokes. This research presents an innovative approach to predicting strokes.

In an alternative medical scenario, Barstugan M. et al. (2020) redirect their attention to the classification of COVID-19 through the analysis of CT [9] images. They underscore the versatility of medical image analysis techniques in addressing diverse medical conditions. Conclusively, Han T. et al. (2020) investigate the utilization of deep learning techniques [3,24], particularly in the segmentation of lung and stroke regions within CT [24] scans. This underscores the adaptability and significance of deep learning in the realm of medical imaging.

Moreover, the field of brain tumor image segmentation has been significantly influenced by a number of noteworthy papers [13]. Notable contributions include works by Dvořák P., Menze B. (2015), Lyksborg M., Puonti O., Agn M., Larsen R. (2015), and Menze B.H., Jakab A., Bauer S., et al. (2014). These publications have been instrumental in the progress of brain tumor image segmentation.

Furthermore, a comprehensive review article by Manish Sirsat, Eduardo Ferme, Joana Camara (2020) offers an extensive overview of the application of machine learning in brain stroke [15,21] detection, providing insights into the broader landscape of this critical medical application. Harish Kamal, Victor Lopez, Sunil A. Sheth's paper (2018) likely delves into the application of machine learning in the context of acute ischemic stroke diagnosis using neuroimaging data, shedding light on the diagnostic potential of these technologies.

The research conducted by Chuloh Kim, Vivienne Zhu, Jihad Obeid, and Leslie Lanert in 2019 appears to revolve around employing natural language processing and machine learning techniques to identify acute ischemic stroke in MRI reports [2,16]. Their work addresses the critical aspect of data interpretation in stroke diagnosis. On the other hand, the paper by R.P. Lakshmi, M.S. Bbu, and V. Vijayalakshmi from 2017 seems to emphasize the utilization of Support Vector Machines (SVM) for voxel-based lesion segmentation in the detection of brain strokes [28], highlighting the significance of SVM in detailed analysis.

Finally, the work conducted by Jae-woo Lee, Hyun-sun Lim, Dong-wook Kim, and Soon-ae Shin may center around methodologies and software related to the analysis of biomedical images [18] within the realm of "Computer Methods and Programs in Biomedicine." Philip A. Wolf, Ralph B. D'Agostino, Albert J. Belanger, and William B. Kannel's investigation likely pertains to assessing stroke risk using data from the Framingham Study, providing valuable insights into the predictive aspects of stroke research [15]. Furthermore, the research by Mn SN, Park SJ, Kim DJ, Subraniyam M, and Lee KS seems to concentrate on developing an algorithm for predicting strokes, utilizing data from a national health insurance database to enhance our understanding of this crucial medical condition.

The problem of detecting and understanding strokes has a significant historical timeline within the field of medicine. Recognized for centuries, strokes [19] have left traces in historical records dating back to ancient civilizations, with ancient Egyptian and Greek medical texts mentioning symptoms indicative of strokes. In a significant historical milestone, the ancient Greek physician Hippocrates, around 460-377 BC, made early observations regarding strokes, referring to a condition he termed "apoplexy," believed to describe strokes as we recognize them today

III. METHODOLOGY

In the field of healthcare, the progress of machine learning algorithms [27] for the identification of brain strokes [7,14,20] is advancing in a rapid pace, allowing people to benefit from life-saving technology but it hasn't been seen it's wide adoption.

```
Informations Of The Dataset :

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10091 entries, 0 to 10090
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   id                 10091 non-null  int64
 1   gender             10091 non-null  object
 2   age                10091 non-null  float64
 3   hypertension       10091 non-null  int64
 4   heart_disease      10091 non-null  int64
 5   ever_married       10091 non-null  object
 6   work_type          10091 non-null  object
 7   Residence_type     10091 non-null  object
 8   avg_glucose_level  10091 non-null  float64
 9   bmi                9890 non-null   float64
 10  smoking_status     10091 non-null  object
 11  stroke             10091 non-null  int64
dtypes: float64(3), int64(4), object(5)
memory usage: 946.2+ KB
None
```

Figure 1. Information of the Dataset

Here is a step-by-step methodology we can follow:

1.Dataset Collection:

The primary sources of data include a huge amount of health records, medical imaging archives, and stroke registries, offering patient information. The process requires extensive preprocessing, including data cleaning, standardization, and feature engineering. The dataset is then divided into training, validation, and test sets to facilitate model development and evaluation. Components like data augmentation, balancing, and privacy considerations play integral roles. Clinician validation ensures the accuracy and quality of the data. Effective data management, continuous data collection, and ongoing algorithm refinement are crucial to maintaining the algorithm's utility and accuracy in clinical practice.

2.Processing:

In the construction of a machine learning algorithm designed for the detection of brain strokes [20], the data processing phase is pivotal. The process involves handling a substantial dataset of 10,000 instances for Random Forest, Naive Bayes, K-nearest Neighbors, SVM, and Logistic Regression. Additionally, 2,500 preprocessed images are utilized for the CNN model. The workflow for data processing encompasses crucial steps such as data collection, integration, preprocessing, feature engineering, labeling, data splitting, data augmentation, dataset balancing, data normalization, and ensuring data security and privacy in compliance with healthcare regulations. The diligently processed data serves as the cornerstone for training and deploying these machine learning algorithms, with the ultimate goal of achieving precise detection of brain strokes. This contributes to improved healthcare outcomes and the well-being of patients.

3.Splitting the Dataset:

In the process of developing a machine learning algorithm [27] dedicated to brain stroke detection [19], the pivotal stage involves the splitting of data. With the incorporation of algorithms like Random Forest, Naive Bayes, K-nearest Neighbors, Support Vector Machine (SVM), Logistic Regression, and Convolutional Neural Network (CNN) [11,12], a significant dataset consisting of 10,000 instances is designated for Random Forest, Naive Bayes, K-nearest Neighbors, SVM, and Logistic Regression. Furthermore, 2,500 preprocessed images are specifically set aside for the CNN model. This dataset is then divided into distinct subsets: a training set, serving as the foundation for model training; a validation set, used for fine-tuning hyperparameters; and a test set, essential for evaluating the algorithms' performance.

4.System Architecture:

In this framework, a substantial dataset of 10,000 instances is allocated for Random Forest, Naive Bayes, K-nearest Neighbors, SVM, and Logistic Regression, while an additional set of 2,500 preprocessed images is reserved for the CNN model training.

In summary, this architecture provides a flexible and comprehensive approach, simplifying the development, deployment, and assessment of various machine learning algorithms for precise and dependable brain stroke detection [21].

Then using the defined split ratios, we divide the shuffled dataset into three subsets:

Training Set:

The training set, comprising 70% of 10,000 instances for algorithms like Naive Bayes, K-nearest Neighbors, SVM, Random Forest, Logistic Regression, and 70% of 2,500 images for the CNN, is the core of the "Machine Learning Algorithm for Brain Stroke Detection," facilitating model development and improving stroke detection accuracy.

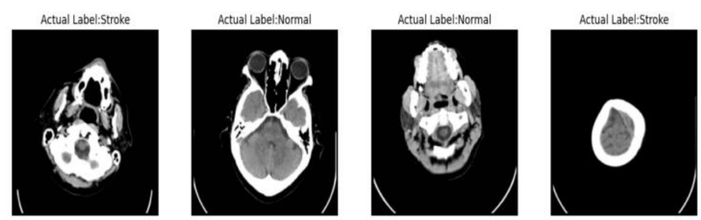

Figure 2. Example of the Training Dataset

Validation Set:

The validation set, comprising 15% of the 10,000 instances (1,500 instances) for Naive Bayes, K-nearest Neighbors, Random Forest, SVM and Logistic Regression, and 15% of the 2,500 images (375 images) for CNN, is pivotal for fine-tuning these algorithms within the "Machine Learning Algorithm for Brain Stroke Detection" system. It facilitates the optimization of model performance, ensuring that the algorithms can effectively generalize to new data, ultimately improving the accuracy of brain stroke detection.

Testing Set:

The testing set, distinct from the training and validation data, is crucial for evaluating the performance of algorithms (Naive Bayes, Random Forest, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, CNN) in brain stroke detection. With 15% of 10,000 instances (1,500 samples) for some algorithms and 15% of 2,500 images (375 images) for CNN, it provides a real-world simulation for accurate assessment. This process aids in refining and enhancing the algorithms' effectiveness in detecting brain strokes.

At last, we had to preserve the split and ensured that the split is consistent across different experiments and iterations. So if we plan to experiment with different models or hyperparameters, we can use the same split for fair comparisons and reproducibility. Then on the second phase, we loaded those pre-processed datasets and took the training dataset and validation dataset for training the model while using the transfer-based learning model with CNN.

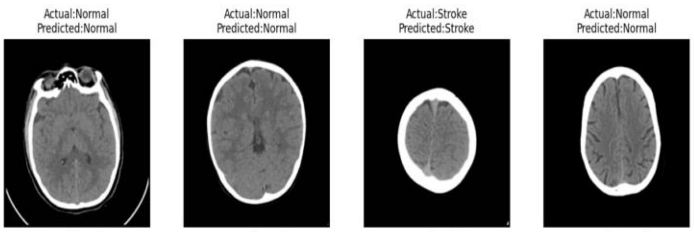

Figure 3. Example of the Test Dataset

Figure 4. Accuracy and Loss graph

Addressing the classification and retrieval problem concerning stroke [20,21] detection entails employing a range of machine learning techniques and algorithms, each offering distinctive approaches and advantages.

Random Forest: Random Forest leverages decentralized training of numerous decision trees [25,26] to solve classification problems. The ultimate classification result is determined by the majority vote of these trees. This ensemble approach contributes to improved model performance, particularly when dealing with complex problems. Random Forest employs multiple Decision Tree datasets and utilizes criteria such as entropy to enhance prediction accuracy. Notably, the Accuracy of Random Forest Classifier is an impressive 98.89%.

```
Confusion Matrix :

[[1884    3]
 [  19   72]]

Classification Report :

              precision    recall  f1-score   support

           0       0.99      1.00      0.99      1887
           1       0.96      0.79      0.87        91

    accuracy                           0.99      1978
   macro avg       0.98      0.89      0.93      1978
weighted avg       0.99      0.99      0.99      1978


The Accuracy of Random Forest Classifier is 98.89 %
```

Figure 5. Confusion Matrix of the Random Forest

Naive Bayes: This is a type of supervised learning, is built upon the Naive Theorem and is widely employed in machine learning. It operates on the premise that the existence of one feature is unrelated to the existence of another feature. Utilizing Bayes' Theorem to calculate conditional probabilities, various versions of Naive Bayes exist, such as Gaussian and Bernoulli Naive Bayes, determined by how class calculations are performed. Notably, the reported accuracy of Bernoulli Naive Bayes is an impressive 94.84%, while the Gaussian variant boasts an accuracy of 87.51%.

```
Confusion Matrix :

[[1870   17]
 [  85    6]]

Classification Report :

              precision    recall  f1-score   support

           0       0.96      0.99      0.97      1887
           1       0.26      0.07      0.11        91

    accuracy                           0.95      1978
   macro avg       0.61      0.53      0.54      1978
weighted avg       0.92      0.95      0.93      1978


The Accuracy of Bernoulli Naive Bayes is 94.84 %
```

Figure 6. Confusion Matrix of the Bernoulli Naive

```
Confusion Matrix :

[[1683  204]
 [  43   48]]

Classification Report :

              precision    recall  f1-score   support

           0       0.98      0.89      0.93      1887
           1       0.19      0.53      0.28        91

    accuracy                           0.88      1978
   macro avg       0.58      0.71      0.61      1978
weighted avg       0.94      0.88      0.90      1978


The Accuracy of Gaussian Naive Bayes is 87.51 %
```

Figure 7. Confusion Matrix of the Gaussian Bayes

K-Nearest Neighbor (KNN): It is identified as a lazy learning algorithm, avoids making explicit assumptions and retains all computations for classification purposes. It categorizes data by assessing the closeness of neighbors in the feature space, utilizing Euclidean distance as the metric. The most effective 'k' value, influencing the classifier's performance, can be determined through a methodical investigation, with your study selecting 'k' as 3. It's noteworthy in this specific instance that the K Nearest Neighbors Classifier demonstrated an impressive Accuracy of 95.75%.

```
Confusion Matrix :

[[1880    7]
 [  77   14]]

Classification Report :

              precision    recall  f1-score   support

           0       0.96      1.00      0.98      1887
           1       0.67      0.15      0.25        91

    accuracy                           0.96      1978
   macro avg       0.81      0.58      0.61      1978
weighted avg       0.95      0.96      0.94      1978


The Accuracy of K Nearest Neighbors Classifier is 95.75 %
```

Figure 8. Confusion Matrix of the K Nearest Neighbors

Support Vector Machine (SVM): This is a supervised learning approach, employs decision planes or hyperplanes to define boundaries for data classification. Its objective is to categorize data by creating a function that assigns each data point to its correct label with minimal error, showcasing its potency in classification tasks. Notably, the Support Vector Machine demonstrates an impressive accuracy of 95.4%, underscoring its efficacy in achieving precise data classification.

```
Confusion Matrix :

[[1887    0]
 [  91    0]]

Classification Report :

              precision    recall  f1-score   support

           0       0.95      1.00      0.98      1887
           1       0.00      0.00      0.00        91

    accuracy                           0.95      1978
   macro avg       0.48      0.50      0.49      1978
weighted avg       0.91      0.95      0.93      1978


The Accuracy of Support Vector Machine is 95.4 %
```

Figure 9. Confusion Matrix of the Support Vector Machine

Logistic Regression: In addition, it's worth noting that the accuracy of Logistic Regression, a prominent member of supervised machine learning algorithms, is remarkably high at 95.4%. Logistic Regression employs multiple independent variables to estimate the probability of a variable, and it is explicitly designed [22] for addressing classification problems, distinguishing it from linear regression, which is tailored to regression tasks. Furthermore, for the analysis of multilinear data, the study mentions the use of ridge regression, a variant of logistic regression.

```
Confusion Matrix :

[[1887    0]
 [  91    0]]

Classification Report :

              precision    recall  f1-score   support

           0       0.95      1.00      0.98      1887
           1       0.00      0.00      0.00        91

    accuracy                           0.95      1978
   macro avg       0.48      0.50      0.49      1978
weighted avg       0.91      0.95      0.93      1978


The Accuracy of Logistic Regression is 95.4 %
```

Figure 10. Confusion Matrix of the Logistic Regression

Convolutional Neural Network (CNN): This is consider a profound advancement in deep learning [5,24], operates by employing convolutional layers to systematically extract intricate features from complex input data. Its unique architecture is tailored to tasks demanding a nuanced understanding of spatial relationships and hierarchical representations, making it exceptionally proficient in domains like image recognition and classification. The notable Accuracy of CNN stands at an impressive 98.6%, a testament to its capacity for achieving highly accurate classifications, thereby solidifying its reputation as a powerful and reliable tool in the realm of data classification. This accuracy rate underscores its utility not only in practical applications but also in research areas where precision and reliability are of paramount importance.

```
Confusion Matrix:
 [[147   0]
 [  2 102]]

Classification Report:
               precision    recall  f1-score   support

           0       0.99      1.00      0.99       147
           1       1.00      0.98      0.99       104

    accuracy                           0.99       251
   macro avg       0.99      0.99      0.99       251
weighted avg       0.99      0.99      0.99       251



The Accuracy of Convolutional Neural Network is 99.2 %
```

Figure 11. Confusion Matrix of the Convolutional Neural Network

IV. RESULT ANANLYSIS AND VALIDATION

The development and validation of a machine learning algorithm [27] for brain stroke detection is a critical aspect of healthcare research. This paper discusses the key steps involved in the development and validation of such an algorithm, focusing on the use of various machine learning algorithms.

Dataset Collection:

The foundation of a successful brain stroke detection algorithm lies in the dataset. A diverse and comprehensive dataset comprising medical images (e.g., MRI or CT scans) with a wide range of stroke [17,23] types and stages must be collected. Collaboration with healthcare institutions and the collection of anonymized patient data is essential to ensure data quality and diversity. A large and balanced dataset is crucial for training and evaluating the performance of the machine learning models.

Data Preprocessing:

Effective machine learning algorithm development relies significantly on data preprocessing. This involves essential tasks like resizing images, reducing noise, and enhancing images to guarantee data quality. Accurate labeling and annotation of the data are essential for effective model training. Furthermore, employing data augmentation techniques such as rotation, flipping, and cropping enhances the dataset, ultimately improving model generalization.

Model Fine-Tuning:

The selection of appropriate machine learning algorithms for brain stroke [7,20] detection is a crucial decision. Each algorithm mentioned has its own strengths and weaknesses. Fine-tuning the model architecture, hyperparameters, regularization techniques, and optimization algorithms is essential to optimize the performance of each algorithm.

Training:

During the training stage, pre-processed data is input into chosen machine learning algorithms. Model parameters undergo iterative adjustments to glean insights from the data. To avoid overfitting, the dataset is usually partitioned into training, validation, and test sets. Evaluating crucial metrics like accuracy, precision, recall, and F1-score throughout the training process aids in gauging model performance.

Model Evaluation:

Following training, each machine learning algorithm is evaluated to determine its performance in detecting brain strokes accurately. Evaluation is typically conducted on a separate test dataset to ensure unbiased results. Metrics like sensitivity, specificity, AUC-ROC, and confusion matrices are used to assess the algorithms' performance.

Deployment:

Once the algorithms have been fine-tuned and successfully evaluated, they can be deployed in a healthcare setting. Deployment involves integrating the algorithms into the clinical workflow, ensuring compatibility with existing systems, and addressing issues related to patient data security and privacy. Regular updates and maintenance are essential for ongoing algorithm performance.

Analysis:

The concluding phase of advancing the algorithm encompasses a thorough evaluation of its functionality within real-world healthcare situations. It is imperative to continually monitor sensitivity, specificity, false positive rate, and false negative rate. Additionally, evaluating the influence of these algorithms on clinical decision-making and patient outcomes is vital. Consistent feedback from healthcare professionals plays a pivotal role in enhancing and fine-tuning the algorithms progressively.

The development and validation of machine learning algorithms for brain stroke [20,23] detection is a complex and multifaceted process. It involves careful attention to dataset collection, data preprocessing, model fine-tuning, training, evaluation, deployment, and continuous analysis. The choice of algorithm plays a significant role in the overall success of the system, and their performance should be rigorously evaluated in real-world healthcare settings to ensure the accurate and timely detection of brain strokes, ultimately leading to improved patient outcomes.

Advantages:

High Accuracy:

Machine learning algorithms, particularly CNN, Random Forest, and SVM, can achieve high accuracy in detecting brain strokes. They can learn complex patterns [6,8] from medical images, making them effective in distinguishing stroke-related anomalies.

Fast Processing:

Algorithms like Logistic Regression, Naive Bayes, and KNN are computationally efficient and can quickly process medical images. This speed can be critical in emergency situations, allowing for timely intervention.

Ensemble Learning:

Random Forest algorithm combines multiple decision trees [25,26], providing robustness against overfitting and improving accuracy. It also handles missing data well, which can be an issue in medical datasets.

Non-linearity Handling (SVM):

Support Vector Machines are effective in handling non-linear relationships in data. This is beneficial when dealing with complex and intricate features present in brain stroke [20] images.

Localized Decision Making (KNN):

KNN considers the local characteristics of data points, making it suitable for analyzing patterns [6,8] in localized regions of the brain, which may be essential for early stroke [7] detection.

Interpretability:

Logistic Regression and Naive Bayes models provide interpretable results, making it easier for healthcare professionals to understand and trust the predictions.

Disadvantages:

Data Dependency:

Machine learning algorithms, especially deep learning [24] models like CNN, are data-dependent. They require large and diverse datasets for training. Obtaining such datasets can be challenging, and the quality of data influences the algorithm's performance.

Overfitting (KNN):

KNN is susceptible to overfitting when the dataset is small or noisy. Proper data preprocessing and feature selection are critical to mitigate this issue.

Hyperparameter Tuning (SVM):

SVM's performance is sensitive to the choice of hyperparameters, and selecting the right parameters can be time-consuming and require expertise.

Risk of Misinterpretation (Random Forest):

While Random Forest provides high accuracy, the ensemble nature of the model can make it challenging to interpret the decision-making process.

Imbalanced Datasets:

Datasets with imbalances, where one class (such as stroke) has noticeably fewer occurrences than the other, may result in biased outcomes. It is essential to manage class imbalance cautiously across all algorithms.

Resource Intensive (CNN):

Convolutional Neural Networks [11,12] demand substantial computational resources, including powerful GPUs and CPUs. This can limit their deployment in resource-constrained environments.

Algorithm Sensitivity:

Variations in image acquisition techniques, noise, and artifacts in medical images can significantly impact the performance of machine learning algorithms, necessitating thorough pre-processing.

In summary, the effectiveness of machine learning algorithms, such as Random Forest, Naive Bayes, Logistic Regression, CNN, KNN, and SVM, in brain stroke detection comes with both advantages and disadvantages. The optimal algorithm choice relies on factors like available data, computational resources, and specific healthcare requirements.

V. CONCLUSION AND FUTURE WORK

Conclusion:

In summary, the utilization of a variety of machine learning algorithms for the detection of brain strokes presents a promising pathway for advancing in healthcare sector. Random Forest and Support Vector Machine (SVM) exhibit strong performance. Naive Bayes algorithm efficiently handles high-dimensional data, while Logistic Regression provides valuable insights into linear relationships. The CNN model, specialized in processing image data, holds exceptional promises in identifying brain stroke patterns [6,8]. However, it is crucial to conduct further refinement and extensive clinical validation before widespread implementation in real-world healthcare settings.

Future Work:

In the realm of brain stroke detection, future endeavors hold the potential for refining and innovating machine learning algorithms. The path ahead entails continuous enhancements in algorithmic accuracy and efficiency, propelled by the incorporation of larger and more diverse datasets to bolster model generalization. Moreover, the exploration of real-time detection capabilities and the integration of emerging technologies like edge computing and wearable devices will be pivotal in advancing the early detection and intervention of strokes [14,28]. Collaborative efforts with healthcare institutions, clinicians, and data scientists will be instrumental in bridging the divide between research and practical clinical application, ultimately diminishing the global impact of stroke-related complications and preserving lives. The forthcoming work in this domain is poised to make significant contributions to the healthcare landscape by providing enhanced diagnostic tools and more effective strategies for stroke prevention and treatment.

Acknowledgment

Authors appreciate and express their gratitude to Er. Monika, Er. Jyoti and Er. Gurleen Kaur for their valuable time and input on discussions, and useful comments on the draft.

References

1. Ozaltin, O.; Coskun, O.; Yeniay, O.; Subasi, A. A Deep Learning Approach for Detecting Stroke from Brain CT Images Using OzNet. Bioengineering 2022, 9, 783.
2. El-Dahshan E.-S.A., Hosny T., Salem A.-B.M. Hybrid intelligent techniques for MRI brain images classification. Digit. Signal Process. 2010;20:433–441. doi: 10.1016/j.dsp.2009.07.002.
3. J.M.Al-Tuwaijari , M.A. Jasim and M.Abdul-Baqi Raheem "Deep Learning Techniques Toward Advancement of Plant Leaf Diseases Detection",2nd Al-Noor International Conference For science and Technology (2NICST2020), Baghdad, Iraq, 2020.
4. Saritha M., Joseph K.P., Mathew A.T. Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. Pattern Recognit. Lett. 2013;34:2151–2156. doi: 10.1016/j.patrec.2013.08.017.
5. Li L., Wei M., Liu B., Atchaneeyasakul K., Zhou F., Pan Z., Kumar S.A., Zhang J.Y., Pu Y., Liebeskind D.S., et al. Deep Learning for Hemorrhagic Lesion Detection and Segmentation on Brain CT Images. IEEE J. Biomed. Health Inform.2020;25:1646–1659.doi: 10.1109/ JBHI .2020. 3028243.
6. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
7. Zhu, G.; Bialkowski, A.; Guo, L.; Mohammed, B.; Abbosh, A. Stroke Classification in Simulated Electromagnetic Imaging Using Graph Approaches. IEEE J. Electromagn. RF Microw. Med. Biol. 2020, 5, 46–53. [CrossRef]
8. Kasabov, N.; Feigin, V.L.; Hou, Z.-G.; Chen, Y.; Liang, L.; Krishnamurthi, R.; Othman, M.; Parmar, P. Evolving spiking neural networks for personalised modelling, classification and prediction of spatio-temporal patterns with a case study on stroke. Neurocomputing 2014, 134, 269–279. [CrossRef]
9. Barstugan, M.; Ozkaya, U.; Ozturk, S. Coronavirus (COVID-19) classification using ct images by machine learning methods. arXiv 2020, arXiv:2003.09424.
10. Han, T.; Nunes, V.X.; Souza, L.F.D.F.; Marques, A.G.; Silva, I.C.L.; Junior, M.A.A.F.; Sun, J.; Filho, P.P.R. Internet of Medical Things—Based on Deep Learning Techniques for Segmentation of Lung and Stroke Regions in CT Scans. IEEE Access 2020, 8, 71117–71135. [CrossRef]
11. Dvořák P., Menze B. International MICCAI Workshop on Medical Computer Vision. Springer; Cham, Switzerland: 2015. Local Structure Prediction with Convolutional Neural Networks for Multimodal Brain Tumor Segmentation; pp. 59–71. [Google Scholar]
12. Lyksborg M., Puonti O., Agn M., Larsen R. 19th Scandinavian Conference on Image Analysis. Springer; Copenhagen, Denmark: 2015. An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation; pp. 201–211. [Google Scholar]
13. Menze B.H., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) IEEE Trans. Med. Imaging. 2014;34:1993–2024. doi: 10.1109/TMI.2014.2377694. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
14. Manish Sirsat Eduardo Ferme, Joana Camara, "Machine Learning for Brain stroke: A Review," Journal of stroke and cerebrovascular disease: the official journal of National Stroke Association (JSTROKECEREBROVASDIS), 20220
15. Harish Kamal, Victor Lopez, Sunil A. Sheth, "Machin e Learning in Acute Ischemic Stroke Neuroimaging," Frontiers in Neurology (FNEUR) 2018.
16. Chuloh Kim, Vivienne Zhu, Jihad Obeid and Leslie Lanert, "Natural Language processing and machine learning algorithm to identify brain MRI reports with acute ischemic stroke, "public Library of Science One (PONE) 2019
17. R.P. Lakshmi, M.S. Bbu and V.Vijayalakshmi "Voxel based lesion segmentation through SVM classifier for effective brain stroke detection," International Conference on wireless Communications, Signal Processing and Networking (WiSPNET), 2017
18. "Computer Methods and Programs in the Biomedical"- Jae-woo Lee, Hyun-sun lim, Dong-wook kim, Soon-ae Shin.
19. "Probability of Stroke: A risk Profile from the Framingham Study"- Philip A. Wolf, MD; Ralph B. D' Agostino, PhD, Albert J.Belanger, MA; and William B. Kannel, MD. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 11 Issue V May 2023- Available at www.ijraset.com © IJRASET: All Rights are Reserved | SJ Impact Facto 1101 r 7.538 | ISRA Journal Impact Factor 7.894 |
20. "Development algorithm for Stroke Prediction: A national Health Insurance Database Study" – Mn SN, Park SJ,Kim DJ, Subraniyam M,Lee KS
21. " Stroke Prediction using Artificial Intelligence"- M.Sheetal Singh, Prakash Choudhary.
22. "Medical Software user interfaces, stroke MD appliction design (IEEE)" Elena Zamsa.
23. "Effective analysis and predictive Model of Stroke Disease using classification methods."-A.Sudha, P.Gayathri, N.Jaisankar.
24. "Deep Learning alohorithm for detection of critical findings in head for CT scan: a retrospective study"- Rohit Ghosh, Sweetha Tanamla, Mustafa Bivaji, Norbert G Campeau, Vasantha Kumar Vanugopal.

25. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics, 1991 May-June; 21(3): 660-674. doi: 10.1109/21.97458.
26. Navada A, Ansari AN, Patil S, Sonkamble BA. Overview of use of decision tree algorithms in machine learning. IEEE Control and System Graduate Research Colloquium, ICSGRC, 2011: 37–42.
27. Rahman MM, Rana MR, Alam NAA, Khan MSI. A web-based heart disease prediction system using machine learning algorithms; 2022 June; 12. 64-80.
28. Tazin T, Alam MN, Dola NN, Bari MS, Bourouis S, Khan M. Stroke Disease Detection and Prediction Using Robust Learning Approaches. Journal of Healthcare Engineering, 2021:1-12. doi: 10.1155/2021/7633381