Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

Multi-Omics Integration with Machine Learning for Early Detection of Ischemic Stroke through Biomarkers Discovery

Mia Md Tofayel Gonee Manik[1]

Abstract

A complicated and diverse neurological condition, ischemic stroke (IS) is marked by a high death rate and substantial long-term impairment. Despite extensive research, reliable biomarkers for the clinical diagnosis and prognosis of ischemic stroke remain elusive, and the underlying molecular pathways remain obscure. This study uses a physiologically informed Convolutional Neural Network (BioCNN) to integrate multi-omics and present a novel method for the early identification and classification of ischemic stroke subtypes. Multi-omics Thirty acute ischemic stroke patients who were hospitalized within twenty-four hours after the beginning of symptoms provided data. Multi-omics profiling included mRNA, miRNA, circRNA, and DNA methylation datasets. After rigorous preprocessing, the integrated into a biomedical knowledge graph to enable graph-based learning. The BioCNN model outperformed models based on individual omics layers in terms of prediction performance. Its accuracy was 97.89%, its F1-score was 96.48%, and its AUC was 95.12%. Comparative analyses also revealed that among single-omics models, mRNA data yielded the best results, highlighting the complementary value of multi-omics integration. These findings emphasize the effectiveness of deep learning frameworks combined with integrated multi-omics data for advancing biomarker discovery and accurate classification of ischemic stroke subtypes, offering promising implications for early diagnosis and personalized treatment strategies.

Keywords: *Biomarkers, Ischemic stroke, Machine learning, Mechanisms, Multi-omics, Systems biology.*

Introduction

Many demographic groups are disproportionately affected by ischaemic stroke, one of the world's primary causes of death and permanent impairment. Although stroke is more likely to happen as people age, it can happen at any time in life. Notably, in 2009, approximately 34% of individuals hospitalized for stroke in United States were younger than 65 years old. There is clear evidence of epidemiological inequalities; for example, the prevalence of stroke-related death is greatest among African Americans, and their risk of having a first stroke is approximately double that of Caucasians [1][2]. Similarly, among females who account for nearly 60% of stroke-related deaths biological and lifestyle factors such as longer lifespan, pregnancy-induced hypertension, contraceptive-induced blood pressure changes, and heightened prevalence of stress, depression, and anxiety contribute to elevated risk [3]. Furthermore, African American females exhibit higher stroke incidence than Caucasian females, driven by increased rates of the CDC reports that high blood pressure, being overweight, and having diabetes [4].

Understanding the complex pathophysiology of ischemic stroke is critical for advancing prevention, diagnosis, and treatment. Recent advances in omics technologies spanning genomics, transcriptomics, proteomics, and metabolomics have significantly enhanced our ability to identify and monitor molecular biomarkers associated with neurovascular diseases [5][6]. These biomarkers are pivotal for improving diagnostic and prognostic accuracy and facilitating the development of targeted therapeutic strategies.

Importantly, in order to decrease mortality and improve patient outcomes, it is crucial to recognize strokes quickly [7]. Biomarkers that enable diagnosis within 60 minutes of stroke onset could potentially offer survival and recovery benefits comparable to mobile stroke units, yet in a more cost-effective manner. Integrating data across various omics layers through systems biology and intermits enables a more comprehensive understanding of stroke mechanisms [8][9]. Coupling this integration with machine learning methodologies allows for the identification of robust, multi-dimensional biomarker signatures, enhancing the early detection of ischemic stroke.

[1] College of Business, Westcliff University, Irvine, CA 92614, USA. Email: m.manik.407@westcliff.edu

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

Motivation and Contributions of the Study

One of the leading causes of death and irreversible neurological damage in the world is still ischaemic stroke. Early and accurate diagnosis plays a pivotal role in mitigating its long-term consequences. However, traditional diagnostic techniques, such as neuroimaging, often fail to detect early molecular alterations that precede clinical manifestations. Recent advances in high-throughput multi-omics technologies have enabled comprehensive profiling of molecular signatures at the genomic, transcriptomic, and epigenomic levels. There is great promise in the combination of different biological data with powerful ML algorithms to discover new biomarkers for the early identification of ischaemic stroke. This study is motivated by the critical need to develop a multi-omics-driven, ML-based framework to enhance diagnostic accuracy and enable timely intervention. The main key contributions are as follows:

A novel framework is proposed that integrates multi-omics data including mRNA, miRNA, circRNA, and DNA methylation for ischemic stroke biomarker discovery.

A BioCNN is developed and trained to classify acute ischemic stroke based on multi-omics profiles.

The study introduces a comprehensive pre-processing pipeline to harmonize and normalize heterogeneous omics data for integrative analysis.

The framework demonstrates robust performance in early stroke detection, assessed utilizing important parameters incorporating recall, accuracy, precision, F1-score, and AUC.

Potential biomarkers are identified, providing important information on the molecular processes behind ischemic stroke and guiding future precision medicine strategies.

Justification and Novelty

The justification and novelty of this study are rooted in its innovative integration of multi-omics data comprising mRNA, miRNA, circRNA, and DNA methylation for the early detection and classification of ischemic stroke using a biologically informed deep learning framework. In contrast to conventional methodologies that primarily rely on single-omics analysis or clinical parameters, this study employs a BioCNN guided by a biomedical knowledge graph to effectively model complex molecular interactions. A more thorough comprehension of the pathophysiological processes behind stroke is made possible by this physiologically contextualized approach. Furthermore, the incorporation of graph-based learning within a unified deep learning architecture marks a novel contribution, enhancing scalability and adaptability to other multifactorial diseases. By advancing a multi-layered, precision-oriented framework, this study sets the foundation for more personalized, timely, and accurate medical interventions in ischemic stroke and broader neurological disease contexts.

Organization of the paper

The organization of this paper is organized as follows: Section II examines relevant research on ischaemic stroke diagnosis via biomarker identification and multi-omics integration. Section III details the methodology employing the BioCNN model for early stroke prediction. Section IV presents the experimental results, including performance comparisons with traditional classifiers such as SVM. Finally, Section V brings the study to a close and talks about potential next research areas.

Literature Review

This section reviews existing ML algorithms applied to multi-omics data integration, particularly for the early detection of ischemic stroke through biomarker discovery. Much of the research focuses on leveraging advanced data-driven techniques to improve diagnostic accuracy and identify critical biomarkers.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

A. K. Subudhi et al. (2018) The stroke lesion was effectively segregated by the approach, and the outcome was assessed using measured metrics that forecast the classifier's accuracy. It fared better than previously published results with a Jaccard index of 0.702, dice coefficient of 0.809, accuracy of 0.835, and sensitivity of 0.804 [10].

Lucas et al. (2018) offered a modification of the well-liked Higher-level characteristics are concatenated in a U-Net architecture, which adds additional skip connections to maximize network propagation. Public perfusion datasets are used to test the proposed approach, which shows much better accuracy (33% less surface distance than U-Net) and resilience, especially for tiny, harder-to-detect stroke lesions [11].

Kodama et al. (2018) gathered HRV data in rats using the MCAO paradigm soon after the commencement of an ischaemic stroke. The model derived from the HRV data of three sham-operated rats was used to construct eleven MCAO-operated and eleven sham-operated animals. The information from the other 19 rats was used to validate it. According to the results of the trial, the recommended method's sensitivity and specificity were 82% and 75%, respectively. [12].

Liu et al. (2018) The suggested Res-FCN obtains the average number of false negative lesions per patient was 1.515, and the average dice coefficient was 0.645 following training and evaluation on a sizable dataset of 212 clinically obtained MRIs. Of the several 2D slice-based segmentation methods, the suggested Res-FCN exhibits a very competitive performance when further tested on a publicly available data set, namely ISLES2015-SISS [13].

Chin et al. (2017) The goal of this study is to use CNN's DL algorithm to create an automated early ischaemic stroke detection system. The system will start picture pre-processing as soon as the brain CT scan is entered in order to exclude any areas that cannot have a stroke. In this study, a CNN module that could identify ischaemic stroke was trained and tested using 256 patch pictures. According to the experimental findings, the suggested method's accuracy exceeds 90% [14].

Giri et al. (2017) its capacity to use the 1DCNN to create a classification algorithm that is able to differentiate between EEG and EOG stroke data and EEG and EOG control data. Batch normalisation is used to speed up our model's training process. With just 200 epochs, 62 person data items, and five assessment repetitions of the leave one out scenario, the average accuracy was 0.86 (F-Score 0.861). This result outperforms all common and shallow classifiers used as a comparison (the best accuracy result was 0.69 and the F-Score was 0.72). Our investigation simply employed 24 manually created features with a straightforward feature extraction procedure [15].

Kumar and Jaya Rama Krishniah (2016) discussed There is discussion of the DT classifier's features. DT classification is used to diagnose three brain diseases: ischaemic stroke, haemorrhage, and haemorrhage and tumour. The DT model performs 91%, 93%, and 95% well in identifying ischaemic stroke, haemorrhage, haematoma, and brain tumour, respectively [16].

Ho et al. (2016) present To estimate four perfusion parameters without requiring explicit deconvolution, a unique bi-CNN was developed. ARMSEs were within 5% of the maximum values, indicating that the bi-CNNs generated accurate approximations. Over 80% of the projected perfusion maps agreed with the ground reality, making them helpful for measuring the amount of salvageable tissue in stroke [17].

Table I compares ML models for early identification of ischaemic stroke and shows that the model performs better than traditional classifiers in terms of accuracy, precision, recall, and F1 score.

Summary on ML-Based Detection Techniques for Early Ischemic Stroke

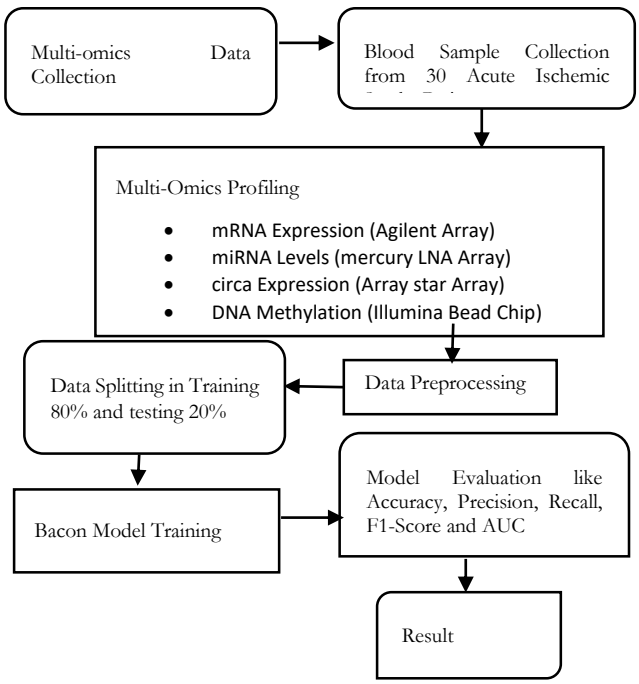
Reference	Methodology	Results	Metrics	Research Gaps	Recommendations
A.K. Subudhi et al. 2018	Segmentation of stroke lesion	Efficient segmentation	Sensitivity: 0.804, Accuracy:	Limited comparison with newer models	Extend evaluation to multiple datasets and

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

	with classifier evaluation		0.835, Dice: 0.809, Jaccard: 0.702		integrate with deep learning methods
Lucas et al. 2018	Extended U-Net with additional skip connections for perfusion data	A 33% reduction in surface distance compared to U-Net and improved lesion segmentation	Improved accuracy, robustness	Performance on large lesions or non-perfusion modalities not discussed	Apply method to multi-modal datasets and compare with 3D segmentation approaches
Kodama et al. 2018	HRV data from MCAO model in rats	Sensitivity: 82%, Specificity: 75%	Sensitivity, Specificity	Small dataset (only 19 rats), limited to animal model	Expand study to human subjects and larger sample sizes for clinical validation
Liu et al. 2018	Res-FCN model trained on 212 clinical MRI and tested on ISLES2015-SISS dataset	Dice: 0.645, FN lesions: 1.515 per subject	Dice coefficient, False Negatives	A lower dice score suggests that segmentation accuracy might be increased	Incorporate 3D CNNs and ensemble methods to enhance segmentation performance
Chin et al. 2017	CNN trained on 256 image patches	Accuracy > 90%	Accuracy	No detailed metrics like sensitivity, specificity, or F-score reported	Provide comprehensive evaluation and validate with real-world clinical cases
Giri et al. 2017	1DCNN for EEG and EOG stroke classification with Batch Normalization	Accuracy: 0.86, F-Score: 0.861	Accuracy, F-Score	Limited to time-series signals; multimodal integration not considered	Combine EEG/EOG with imaging for better diagnosis; test scalability to large datasets
Kumar & Jaya Rama Krishniah, 2016	Decision Tree classifier for brain disease diagnosis	Stroke: 91%, Hemorrhage: 93%, Tumor: 95%	Accuracy	Classical ML model; lacks deep learning comparison	Evaluate using deep learning techniques; test model robustness across diverse imaging modalities
Ho et al. 2016	Bi-input CNN to estimate perfusion parameters without deconvolution	ARMSE ≤ 5% of max, >80% agreement with ground truth	ARMSE, Agreement with ground truth	Limited discussion on lesion localization or segmentation performance	Integrate bi-CNN with lesion detection models for holistic stroke analysis

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

Methodology



Flowchart of Framework for Early Detection of Ischemic Stroke

The proposed framework integrates multi-omics data with machine learning techniques to facilitate the early detection of AIS through biomarker discovery. As illustrated in Figure 1, the study begins with multi-omics data collection involving blood samples from 30 individuals diagnosed with AIS. Subsequently, multi-omics profiling is performed, including mRNA expression (Agilent Array), miRNA levels (mercury LNA Array), circa expression (Array star Array), and DNA methylation (Illumina Bead Chip). Pre-processing procedures are used to clean and normalize the data once it has been acquired, guaranteeing interoperability across omics layers. The dataset is then divided into subgroups for testing (20%) and training (80%). A biologically inspired convolutional neural network (BioCNN) is trained using the processed data to classify and predict AIS. Using common classification measures including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), the model's performance is thoroughly assessed following training. The results highlight potential multi-omics biomarkers that contribute significantly to the early diagnosis of ischemic stroke, offering promising insights into precision medicine.

Data Collection and Analysis

The multi-omics information gathered from 30 acute ischaemic stroke patients who were hospitalized to the Hospital of Navarra in Spain within twenty-four hours of the beginning of their symptoms. These patients were classified into three etiological subgroups using the TOAST classification system: atherothrombotic, cardioembolic, and unknown. Blood samples were taken from each patient and analysed using four high-throughput omics platforms The mRNA expression was profiled using the Agilent Surprint G3 Human Gene Expression 8 × 60K v3 array. miRNA levels were measured using the mercury LNA microRNA Array 7th Generation. Circa expression was evaluated using the Array star Human circRNA Array V2. The Illumina Human Methylation EPIC Bead Chip array was used to analyse genome-wide DNA methylation.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

Data Preprocessing

Data preparation is an umbrella term that describes a set of processes that prepare the data for the algorithms to be processed. The data that are available from various sources may contain irregularities that can negatively impact the performance of the algorithm. In this study, the raw omics data underwent extensive pre-processing to ensure quality and consistency. Datasets, background correction and normalization procedures were applied. Low-quality samples and outliers, five mRNA samples that failed quality control, were excluded. Methylation data pre-processing included filtering out probes on sex chromosomes, those overlapping common SNPs, cross-reactive probes, and probes failing detection p-value or bead count thresholds. Expression probes were mapped to gene identifiers using resources such as GEO, Ensemble, miRbase, Mirta Base, and CircInteractome. Finally, the cleaned datasets were integrated into a biomedical knowledge graph to capture interactions between different molecular entities across omics layers, forming the foundation for subsequent graph-based analysis.

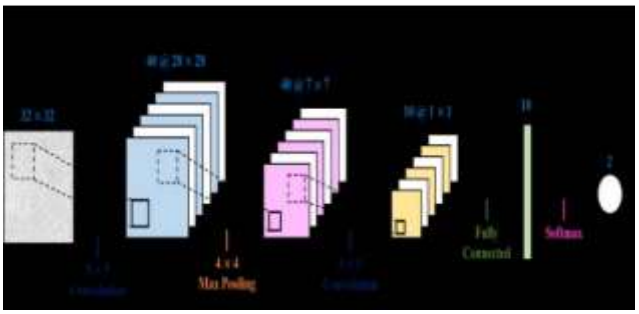
Data Splitting

The training and testing set of the dataset were divided at a ratio of 80% and 20%, respectively. The dataset that ML algorithms and ensemble algorithms must use is known as the training set. The testing dataset is the multi-omics dataset that shows how accurate the model.

Convolutional Neural Network (BioCNN) for Early Detection of Ischemic Stroke

ML model improves early disease detection by integrating complex multi-omics data to identify critical biomarkers, therefore improving clinical judgement and diagnostic precision. This research evaluated the effectiveness of a Convolutional Neural Network-based model (BioCNN) in the early detection of ischaemic stroke, demonstrating superior accuracy and robustness compared to traditional classifiers, highlighting its potential in biomarker-driven precision medicine.

CNN are a type of feed-forward, DANN that has been effectively used for visual imagery analysis. CNNs employ a multilayer perceptron variant that requires little pre-processing [18]. CNN uses a modified version of the backpropagation learning technique to determine the error between the predicted and actual outputs. Gradient Descent is then used to discover the local best solution for adjusting the weights of each layer. A CNN is made up of several hidden layers in addition to an input and output layer. Three kinds of hidden layers exist: completely connected, pooling, and convolutional [14]. One pooling layer, one fully connected layer, and two convolutional layers make up the CNN architecture's five layers. The CNN architecture is presented in Figure 2.



The CNN Architecture has Five Layers

Convolutional layer

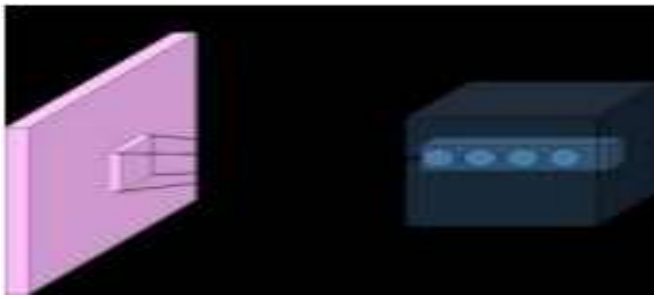
The input image's many attributes are captured by the convolutional layer. Convolution operations should be applied to the input once convolutional layers are finished, and the results should be sent to the next layer. Only the receptive field of each convolutional neurone may be used for data processing. Generalization is enhanced and the number of free parameters is decreased by the convolution procedure.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

Put another way, it uses the backpropagation technique to overcome the disappearing or exploding issues that arise while training a traditional multi-layer neural network, which consists of many layers. The interaction between the convolutional layer and the picture is shown in Figure 3. Each convolutional layer uses ReLU nonlinearity as its activation function, and the input patch image is 32 x 32. In every convolutional layer, the filter size is 5 x 5. Equation (1) defines the well-known logistic sigmoid function, which performs worse than the ReLU:

$$f(x) = x^{+} = \max(0, x) \qquad (1)$$

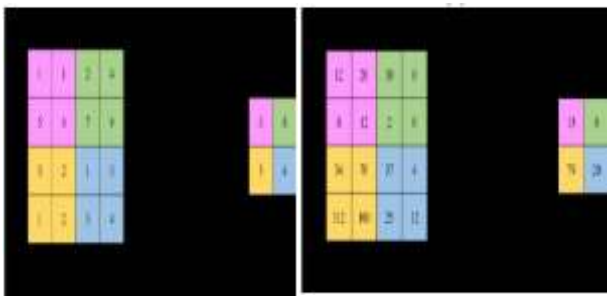
where x represents input data.



The Operation between the Image and the Convolutional Layer

Pooling Layer

In order to minimize the issue of overfitting, The pooling layer's primary function is down-sampling, which entails reducing the quantity of parameters and data. The pooling layer in the next layer combines the outputs of neurone clusters in the previous layer to form a single neurone. Figure 4 illustrates the max pooling method in (a) and the average pooling method in (b). Each pooling layer's maximum pooling and filter size is 4x4.



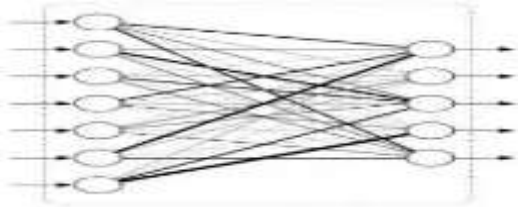
(a) (b)

(A) The Way of Max Pooling, and (B) The Way of Average Pooling.

Fully Connected Layer

The neural network's fully connected layers do the high-level reasoning following a number of convolutional and pooling layers. Layers that are fully linked link each neurone in one layer to every other layer's neurone. In normal neural networks, every neurone in the fully connected layer is related to every activation in the layer above. The fully connected layer's diagram is shown in Figure 5.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800



The Fully Connected Layer Diagram

Model Evaluation

The performance measurements utilized in performance indicators include F1 score, AUC, recall, accuracy, and precision. These can be classified as FP or FN, but if people are right, they can be classified as TP or TN. The most popular measures are explained here.

Accuracy: The accuracy indicates the total number of successfully identified cases out of all the instances. To determine accuracy, use the formula below Equation (2).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

Precision: Precision is measured by dividing all expected positive observations by the number of properly predicted observations, which is formulated in Equation (3).

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall: Recall is defined as the percentage of total relevant results that the algorithm correctly detects and is expressed in Equation (4).

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1-score: The F1 score is calculated by taking the harmonic mean of recall and accuracy. Equation (5) formulates the maximum F score of 1, which denotes flawless accuracy and recall score.

$$F1\ Score = \frac{2\ (Precision \times Recall)}{Precision \times Recall} \quad (5)$$

Area under curve (AUC): The models' behavior under various conditions is displayed by the AUC. The formula below Equation (6) may be used to calculate the AUC.

$$AUC = \frac{\sum R_i(I_p) - I_p((I_p+1)/2)}{I_p+I_n} \quad (6)$$

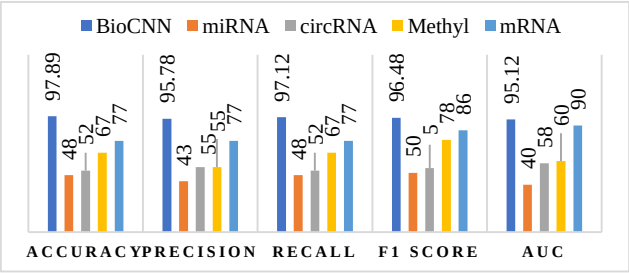
Result Analysis and Discussion

The experimental method aims to enhance the early detection and classification of ischemic stroke by leveraging multi-omics data integration and advanced ML techniques. Which combines real-world patient profiles with synthetic attack patterns for robustness, implemented the PyTorch Geometric (version 2.1.0) BioCNN model. Using the Adam optimiser, the model was trained on a single NVIDIA H100 GPU with 80 GB of RAM. Following training, the finished patient embeddings were taken out for classification exercises. As shown in Table II, the BioCNN model significantly outperformed single-omics approaches such as miRNA, circRNA, methylation (Methyl), and mRNA across all evaluation metrics.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

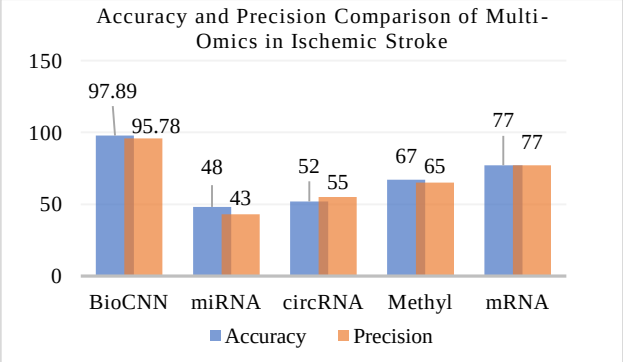
Evaluation metrics for the patient etiology classification task.

Model	Accuracy	Precision	Recall	F1 Score	AUC
BioCNN	97.89	95.78	97.12	96.48	95.12
miRNA	48	43	48	50	40
circRNA	52	55	52	54	58
Methyl	67	65	67	78	60
mRNA	77	77	77	86	90



Performance Metrics of Multi-Omics Biomarkers in Ischemic Stroke Detection using BioCNN

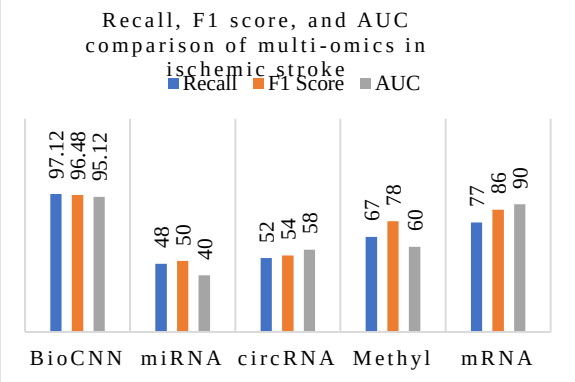
The suggested BioCNN model's performance comparison versus the individual multi-omics biomarkers miRNA, circRNA, methylation, and mRNA for ischaemic stroke diagnosis is shown in Figure 6. Assessment metrics include AUC, F1-score, recall, accuracy, and precision. BioCNN significantly outperforms all individual biomarker models, achieving the greatest results in every metric: 95.12% AUC, 96.48% F1-score, 97.89% accuracy, 95.78% precision, and 97.12% recall. In contrast, individual omics data show lower performance, with miRNA and circRNA yielding particularly reduced precision and recall. These results highlight the effectiveness of BioCNN in integrating multi-omics data for superior diagnostic accuracy in ischemic stroke detection.



Comparative Analysis of Accuracy and Precision Across Multi-Omics Modalities Using BioCNN

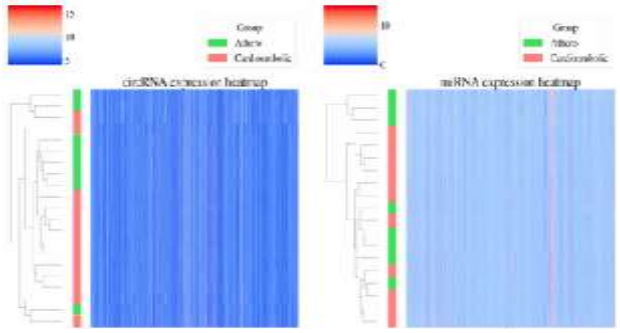
Figure 7 shows the comparative performance of different omics data modalities miRNA, circRNA, Methylation, and mRNA against the integrated model BioCNN in terms of precision and accuracy for ischaemic stroke early detection. The BioCNN model, which integrates multiple omics layers, significantly outperforms individual data types, achieving the highest accuracy 95.78% and precision 97.89%. In contrast, individual omics such as miRNA and circRNA exhibit lower performance metrics, emphasizing the advantage of multi-omics integration. This result highlights the effectiveness of the BioCNN model in improving stroke biomarker discovery through ML.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800



Comparative Analysis of Recall, F1 Score, and AUC Across Multi-Omics Modalities

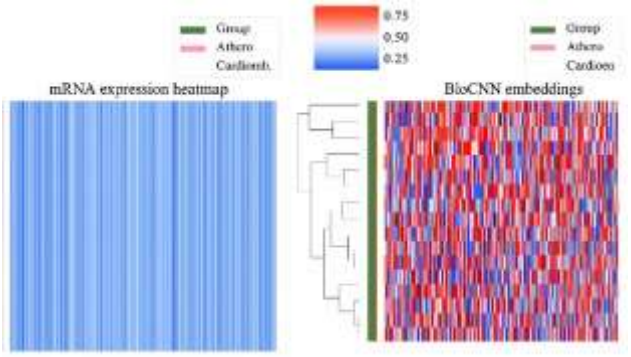
A comparison of recall, F1-score, and AUC for five different data modalities is shown in Figure. 8 BioCNN (multi-omics integrated), miRNA, circRNA, Methylation, and mRNA. The BioCNN model achieves the highest performance across all three metrics, reporting a 95.12% AUC, 97.12% Recall, and 96.48% F1-Score. Among the individual omics-based models, mRNA exhibits comparatively strong performance, with a Recall of 77% and AUC of 90%, while miRNA yields the lowest values. These results highlight how multi-omics integration with DL might improve biomarker identification and facilitate ischaemic stroke early diagnosis.



Circrna and miRNA Expression Heatmaps for Ischemic Stroke Subtype Classification

Figure 9 presents the expression heatmaps of circa and miRNA across two ischemic stroke subtypes Atherothrombotic (green) and Cardioembolic (pink). Hierarchical clustering highlights group-specific expression variations, with both omics' profiles displaying limited but distinct subgroup separation. The expression intensity is represented by the blue-red gradient, ranging from low to high expression levels. Despite visible differences in expression patterns, classification performance using miRNA and circRNA data alone remains suboptimal. The accuracy and AUC of circRNA were 52 and 0.58, respectively, whereas miRNA produced an even lower accuracy of 48 and 0.40. These findings emphasize the necessity of integrating multiple omics layers to enhance early ischemic stroke detection through machine learning.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800



Mrna Expression and BioCNN Embedding Heatmaps for Ischemic Stroke Subtype

Figure 10 compares mRNA expression heatmaps with BioCNN-generated embeddings across ischemic stroke subtypes Atherothrombotic (green) and Cardioembolic (pink). The mRNA heatmap displays moderate subtype separation, aligning with its classification performance an accuracy of 0.77 and AUC of 0.90 demonstrating its diagnostic relevance. In contrast, the BioCNN embeddings capture integrated multi-omics patterns, highlighting pronounced subtype distinction through enhanced feature abstraction. This visual differentiation reflects the superior classification capabilities of the CNN-based model. These findings underscore the effectiveness of DL-based multi-omics integration in early stroke subtype classification.

Comparison and Discussion

The performance of the proposed BioCNN model and the SVM model for multi-omics data-based ischaemic stroke diagnosis is assessed in Table III. The BioCNN model achieves superior results across all evaluation metrics, has a 96.48% F1-score, 97.89% accuracy, 95.78% precision, and 97.12% recall. In contrast, the SVM model reports consistent values of 92% for F1-score, recall, and precision, as well as a somewhat reduced accuracy of 94%. These results demonstrate the enhanced predictive capability of the BioCNN model when applied to integrated multi-omics datasets for stroke detection.

Performance Comparison of BioCNN and SVM Models for Ischemic Stroke Detection

Matrix	**BioCNN**	**SVM** [19]
Accuracy	97.89	94
Precision	95.78	92
Recall	97.12	92
F1 Score	96.48	92

The proposed advantage of this works its comprehensive evaluation and demonstration of the BioCNN model's better performance when contrasted with more conventional ML techniques like SVM for early detection of ischemic stroke through multi-omics biomarker integration. This approach highlights the enhanced sensitivity, robustness, and generalization capabilities of DL models when applied to complex biological datasets. By leveraging advanced CNN architectures, the work underscores the potential of integrating multi-omics data to develop more accurate and scalable tools for early ischemic stroke diagnosis, offering significant improvements over conventional classification techniques.

Conclusion and Future Scope

Stroke is a major contributor to mortality and morbidity, placing a heavy cost on modern healthcare systems. Because of its high mortality and long-term impairment rates, ischemic stroke continues to be a significant challenge to international healthcare systems. This study presents an effective approach that integrates multi-omics data with a biologically inspired deep learning model BioCNN for the early

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

identification and categorization of stroke. With an accuracy of 97.89%, the suggested model performed better than the others, outperforming both single-omics frameworks and conventional ML classifiers such as SVM. These results highlight how crucial multi-omics integration is to understanding the intricate molecular landscape of ischemic stroke, thereby facilitating the identification of robust biomarkers and enabling precision diagnostics.

Future work will aim to extend this framework by validating the strategy to guarantee generalizability using bigger and more varied patient cohorts. Additional omics layers, including proteomics and metabolomics, will be incorporated to enrich the molecular context and improve predictive power. Moreover, emphasis will be placed on developing interpretable DL architectures and integrating them into real-time clinical decision support systems to enhance usability and adoption in clinical settings. Such advancements are expected to contribute significantly to the reduction of the worldwide burden of ischemic stroke by early detection and individualized treatment.

References

[1] C. A. Jackson and G. D. Mishra, "Depression and Risk of Stroke in Midaged Women," Stroke, vol. 44, no. 6, pp. 1555–1560, Jun. 2013, doi: 10.1161/STROKEAHA.113.001147.

[2] V. Kolluri, "Machine Learning in Managing Healthcare Supply Chains: How Machine Learning Optimizes Supply Chains, Ensuring the Timely Availability of Medical Supplies," J. Emerg. Technol. Innov. Res., vol. 3, no. 6, 2016.

[3] S. Sacco et al., "Hormonal contraceptives and risk of ischemic stroke in women with migraine: a consensus statement from the European Headache Federation (EHF) and the European Society of Contraception and Reproductive Health (ESC)," J. Headache Pain, 2017, doi: 10.1186/s10194-017-0815-1.

[4] L. Ramiro, A. Simats, T. García-Berrocoso, and J. Montaner, "Inflammatory molecules might become both biomarkers and therapeutic targets for stroke management," 2018. doi: 10.1177/1756286418789340.

[5] A. Kunz et al., "Effects of Ultraearly Intravenous Thrombolysis on Outcomes in Ischemic Stroke," Circulation, vol. 135, no. 18, pp. 1765–1767, May 2017, doi: 10.1161/CIRCULATIONAHA.117.027693.

[6] A. O. Badheka et al., "ST-T wave abnormality in lead avr and reclassification of cardiovascular risk (from the national health and nutrition examination survey-III)," Am. J. Cardiol., 2013, doi: 10.1016/j.amjcard.2013.04.058.

[7] R. Senn, M. S. V. Elkind, J. Montaner, M. Christ-Crain, and M. Katan, "Potential Role of Blood Biomarkers in the Management of Nontraumatic Intracerebral Hemorrhage," Cerebrovasc. Dis., vol. 38, no. 6, pp. 395–409, 2014, doi: 10.1159/000366470.

[8] J. Montaner et al., "Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke," 2020. doi: 10.1038/s41582-020-0350-6.

[9] V. Kolluri, "An Innovative Study Exploring Revolutionizing Healthcare With AI: Personalized Medicine: Predictive Diagnostic Techniques and Individualized Treatment," JETIR - Int. J. Emerg. Technol. Innov. Res. (www. jetir. org| UGC issn Approv. ISSN, vol. 3, no. 11, pp. 2349–5162, 2016.

[10] A. K. Subudhi, S. S. Jena, M. Mohanty, and S. K. Sabut, "Computational Intelligence Approach for Predicting Ischemic Stroke using Brain MRI," in Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018, 2018. doi: 10.1109/ICICCT.2018.8473213.

[11] C. Lucas, A. Kemmling, A. M. Mamlouk, and M. P. Heinrich, "Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images," in Proceedings - International Symposium on Biomedical Imaging, 2018. doi: 10.1109/ISBI.2018.8363767.

[12] T. Kodama et al., "Ischemic Stroke Detection by Analyzing Heart Rate Variability in Rat Middle Cerebral Artery Occlusion Model," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 6, Jun. 2018, doi: 10.1109/TNSRE.2018.2834554.

[13] Z. Liu, C. Cao, S. Ding, Z. Liu, T. Han, and S. Liu, "Towards clinical diagnosis: Automated stroke lesion segmentation on multi-spectral MR image using convolutional neural network," IEEE Access, 2018, doi: 10.1109/ACCESS.2018.2872939.

[14] C. L. Chin et al., "An automated early ischemic stroke detection system using CNN deep learning algorithm," in Proceedings - 2017 IEEE 8th International Conference on Awareness Science and Technology, iCAST 2017, 2017. doi: 10.1109/ICAwST.2017.8256481.

[15] E. P. Giri, M. I. Fanany, A. M. Arymurthy, and S. K. Wijaya, "Ischemic stroke identification based on EEG and EOG using ID convolutional neural network and batch normalization," in 2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016, 2017. doi: 10.1109/ICACSIS.2016.7872780.

[16] D. V. Kumar and V. V. Jaya Rama Krishniah, "An automated framework for stroke and hemorrhage detection using decision tree classifier," in Proceedings of the International Conference on Communication and Electronics Systems, ICCES 2016, 2016. doi: 10.1109/CESYS.2016.7889861.

[17] K. C. Ho, F. Scalzo, K. V. Sarma, S. El-Saden, and C. W. Arnold, "A temporal deep learning approach for MR perfusion parameter estimation in stroke," in Proceedings - International Conference on Pattern Recognition, 2016. doi: 10.1109/ICPR.2016.7899819.

[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, 1998, doi: 10.1109/5.726791.

Journal of Ecohumanism
2023
Volume: 2, No: 2, pp. 175 – 187
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v2i2.6800

[19] A. Subudhi, U. R. Acharya, M. Dash, S. Jena, and S. Sabut, "Automated approach for detection of ischemic stroke using Delaunay Triangulation in brain MRI images," Comput. Biol. Med., vol. 103, pp. 116–129, Dec. 2018, doi: 10.1016/j.compbiomed.2018.10.016.