IRSTI 28.23.37

10.51889/2959-5894.2024.88.4.018

A.T. Tursynova[1*], B.S. Omarov[1]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[*]*e-mail: azhar.tursynova1@gmail.com*

COMPARATIVE ANALYSIS OF THE ENSEMBLE MODEL FOR STROKE CLASSIFICATION

Abstract

This study investigates the efficacy of an ensemble model in classifying stroke images, combining CNN (Convolutional Neural Network), EfficientNetB7, and DenseNet201 architectures. Utilizing a dataset of 2,501 black-and-white images from the Kaggle stroke dataset, the research addresses the challenges posed by limited data and explores data augmentation techniques to improve model performance. The ensemble model's performance is compared against individual models such as MobileNetV2, EfficientNetB0, ResNet50, and DenseNet201. Results demonstrate that, while the ensemble model shows potential, its accuracy does not significantly exceed that of the top-performing standalone models, highlighting the need for larger datasets and more sophisticated ensemble techniques to enhance reliability. This work provides insights into the application of ensemble learning for stroke classification, paving the way for advancements in AI-driven stroke diagnostics.

Keywords: stroke classification, ensemble learning, CNN, EfficientNetB7, DenseNet201, data augmentation, medical image analysis, artificial intelligence, deep learning, brain imaging.

А.Т. Турсынова[1], Б.С. Омаров[1]
[1]Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛИ АНСАМБЛЯ ДЛЯ КЛАССИФИКАЦИИ ИНСУЛЬТА

Аннотация

В данном исследовании изучается эффективность ансамблевой модели для классификации изображений инсульта, объединяющей архитектуры CNN (Convolutional Neural Network), EfficientNetB7 и DenseNet201. Используя набор данных из 2501 черно-белого изображения из открытого набора данных Kaggle для инсульта, исследование направлено на преодоление проблем, связанных с ограниченным объемом данных, и включает применение методов аугментации для улучшения производительности модели. Производительность ансамблевой модели сравнивается с отдельными моделями, такими как MobileNetV2, EfficientNetB0, ResNet50 и DenseNet201. Результаты показывают, что, хотя ансамблевая модель демонстрирует потенциал, её точность незначительно превышает показатели лучших отдельных моделей, что подчеркивает необходимость использования более крупных наборов данных и более сложных методов ансамблирования для повышения надежности. Это исследование предоставляет ценные данные о применении ансамблевого обучения для классификации инсульта, открывая перспективы для развития диагностики инсульта на основе ИИ.

Ключевые слова: классификация инсульта, ансамблевое обучение, CNN, EfficientNetB7, DenseNet201, аугментация данных, анализ медицинских изображений, искусственный интеллект, глубокое обучение, визуализация мозга.

А.Т. Турсынова[1], Б.С. Омаров[1]
[1]әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

ИНСУЛЬТТІ ЖІКТЕУГЕ АРНАЛҒАН АНСАМБЛЬ МОДЕЛІН САЛЫСТЫРМАЛЫ ТАЛДАУ

Аңдатпа

Осы зерттеуде CNN (Convolutional Neural Network), EfficientNetB7 және DenseNet201 архитектураларын біріктіретін ансамбльдік модельдің инсульт суреттерін классификациялаудағы тиімділігі қарастырылады. Kaggle ашық инсульт деректер жинағынан алынған 2501 қара-ақ суреттер жиынтығын пайдалана отырып, зерттеу деректердің шектеулі көлемімен байланысты мәселелерді шешуге бағытталған және модельдің өнімділігін арттыру үшін аугментация әдістерін қолдануды қамтиды. Ансамбльдік модельдің өнімділігі MobileNetV2, EfficientNetB0, ResNet50 және DenseNet201 сияқты жеке модельдермен салыстырылады. Нәтижелер көрсеткендей, ансамбльдік модель әлеуетке ие болғанымен, оның дәлдігі үздік жеке модельдердің көрсеткіштерінен аз ғана жоғары, бұл сенімділікті арттыру үшін үлкен деректер жиынтықтарын және күрделірек ансамбль әдістерін қолданудың қажеттілігін көрсетеді. Бұл зерттеу инсультті классификациялау үшін ансамбльдік оқытуды қолдану туралы құнды ақпарат береді, жасанды интеллект негізіндегі инсульт диагностикасын дамытуға перспективалар ашады.

Түйін сөздер: инсультті классификациялау, ансамбльдік оқыту, CNN, EfficientNetB7, DenseNet201, деректерді аугментациялау, медициналық суреттерді талдау, жасанды интеллект, терең оқыту, ми визуализациясы.

Main provisions

This study looks at how well an ensemble model works for classifying stroke images, combining CNN, EfficientNetB7, and DenseNet201. With a limited dataset, the research shows the ensemble model only gives a small boost in accuracy compared to single models like MobileNetV2 and EfficientNetB0. Using data augmentation techniques made the training set more diverse and improved overall model performance; still, the ensemble method did not offer a big advantage. These results stress the importance of having bigger and more varied datasets, along with better ensemble methods, to see real gains in classification accuracy. This research gives useful insights into using ensemble learning for stroke diagnosis, opening doors for future progress in AI-based medical diagnostics.

Introduction

Stroke is a major health issue around the world, being a top cause of death and lasting disability. The World Health Organization (WHO) reports that about 15 million people have a stroke every year, with 5 million ending in death and another 5 million facing lasting disabilities [1]. The serious effects of stroke affect not just the individuals who survive but also create heavy costs for healthcare systems globally. Quick and correct identification of stroke is key to lowering the death rate and the long-term issues related to it, as it allows for timely treatment [2, 3]. This is especially significant as ischemic and hemorrhagic strokes are treated differently. Neuroimaging studies, and especially CT, are crucial for stroke diagnosis as they enable neurotherapists to quickly evaluate the patients' and/or their brain images. Although MRI is suggested to have superior imaging quality, CT is more useful in emergency cases because it is readily available, quick to perform, and inexpensive [4].

AI and deep learning, in particular, have emerged as technology that can change the game in medicine, enabling improvements in the automated diagnosis over the past few years. Strokedetection is one of the brain imaging classification tasks that have implemented CNNs heavily [5]. CNNs learn from labeled examples of different classes many compounds and features, which leads to sufficiently high levels of output accuracy that can be comparable to those achieved by humans in certain areas. Nevertheless, architectures such as MobileNetV2, EfficientNetB0, ResNet50 and DenseNet201 have some disadvantages [6]. These shortcomings may be more pronounced in the case of small data sets characteristic of medical imaging due to data size and confidentiality issues [6, 7].

Ensemble learning methods have emerged in response to the single-minded models by incorporating the best of the several models [8] and are able to improve accuracy and robustness. The use of ensemble models in the classification of medical images has been encouraging because models with different learning architectures can use them to make improved predictions. This study aims to evaluate the performance of an ensemble model comprising of CNN, EfficientNetB7 and DenseNet201 for stroke classification, arguing that such a combination could perform better than any of the architectures in isolation. In this study we compare this ensemble model with mobile v2, efficientnetb0, resnet50, densenet201 standalone networks to determine whether such ensemble methods provide direct improvement in classification performance as seen in [8].

This research has two primary goals: the first one is about determining whether an ensemble approach can enhance stroke image classification on a small dataset, and the second is concerned with the extent to which the ensemble performs better than the individual models. It is anticipated that the results of this research will show how stroke diagnostics can benefit from the use of ensemble learning and foster further developments in the AI-supportive systems aimed at improving the performance of health practitioners through more accurate and time sensitive diagnosis. In view of the market demands for better instruments towards clinical practice especially in less well-equipped societies, such work makes a contribution towards further development of AI in medical imaging [9].

Literature Review

Stroke remains an issue, for health and is the second most common cause of death and a major contributor to long term disability on a worldwide scale. It is crucial to detect strokes accurately to enhance outcomes through timely intervention since ischemic and hemorrhagic strokes necessitate different treatment strategies. Diagnosing strokes typically begins with neuro imaging techniques, like computed tomography (CT scan) and magnetic resonance imaging (MRI) which are commonly utilized in practice. In emergency situations, like acute care settings where time's crucial and availability is key doctors usually opt for CT scans because they are fast and widely accessible. On the hand MRI scans offer image quality and resolution enhancements albeit at a higher price and limited availability.

Traditional methods, for analyzing stroke images [4] heavily depend on radiologists expertise – a time consuming and error prone process that has sparked an interest in intelligence (AI) deep learning (DL). The use of Convolutional Neural Networks (CNNs) a DL architecture tailored for image analysis tasks like medical image classification for stroke detection and categorization – has shown potential, in providing automated and accurate solutions. Convolutional Neural Networks (CNN) acquire knowledge from collections of categorized images by identifying intricate visual attributes without the need, for extensive manual input; hence they are well suited for tasks such, as stroke categorization.

Research, on learning models for categorizing strokes primarily focuses on neural networks (CNN) and their variations. CNN technology is commonly used to analyze CT and MRI scans to differentiate between hemorrhagic strokes with accuracy. Although CNN models excel in this task on datasets in medical imaging applications due to their effectiveness in extracting features from brain images they encounter challenges when dealing with smaller datasets due to privacy concerns and limited accessibility. For example a study referenced as [1] showcases the capability of CNN models in extracting features, from brain images. Also points out the difficulty they face in generalizing results when trained on limited data. Moreover because MRI data offers tissue details algorithms trained on MRI datasets typically exhibit precision, in detecting subtle stroke related variations when compared to models based on CT scans. However in situations CT scans still present, as the alternative.

In order to improve the precision of diagnoses and address the constraints of using CNN structures researchers have been exploring alternative DL models such, as ResNet, EfficientNet and DenseNet. For example EfficientNet effectively adjusts network parameters to capture image details with less computational effort. On the hand DenseNet has been found to enhance the flow of information, across layers, which proves useful in capturing structural features within regions affected by stroke. Although these models have shown success individually they may encounter performance challenges in situations. When the training dataset lacks diversity.

Ensemble learning has demonstrated outcomes in bolstering the effectiveness of stroke classification systems by merging models to enhance predictive accuracy collectively. In this approach each model lends its patterns to offset model shortcomings and improve the systems overall efficiency. Researchers have investigated combining models, like CNN, EfficientNetB7 Densenet201 to elevate classification precision by leveraging the strengths of each architecture. For example a study analyzed a combination of MRI data, with MobileNet V2. Achieved promising results in identifying signs of a stroke, in the brain.

Additionally incorporating techniques, like information and feature importance ranking has enhanced the effectiveness of models in categorizing strokes. Research has demonstrated that these combined systems not enhance classification accuracy metrics such as sensitivity and specificity but also make it easier to understand the model by emphasizing the crucial image characteristics. This strategy has shown success when employing gradient boosting algorithms such as LightGBm or CatBoost in conjunction, with CNN result.

In the field of stroke image analysis and segmentation models working together have made an impact far. Identifying the areas, in the brain is crucial for classifying strokes and planning treatment. A notable event in this domain is the Ischemic Stroke Lesions Segmentation (ISLES) challenge [10] which has played a role, in developing and evaluating deep learning algorithms for segmenting stroke lesions. The ISLES 2018 challenge especially has significantly contributed to improving segmentation techniques by providing a dataset containing types of CT and MRI images. Having access, to these datasets has allowed researchers to enhance models for segmentation. This is important, for recognizing tissue in cases of strokes [11].

The effectiveness of division and categorization models, in medical image tasks demonstrates how ensemble techniques can be flexible and versatile in scenarios, within the field of healthcare technology. One example is the Stroke Unet network. An architecture based on the Internet of Medical Things (IoMT) specifically crafted for stroke analysis [12]. It merges UNet with layers customized for high resolution CT scans to enable instant data processing that aids clinical decision making processes. This combined method takes advantage of UNets proficiency in segmentation while also utilizing IoMT to integrate with clinical protocols. By integrating a combination of models into the Internet of Medical Things (IoMT) medical professionals can improve the availability of data to allow for monitoring and analysis, in acute stroke care scenarios that support treatment responses.

Although ensemble models have benefits to offer in fields including medicine, they also come with their own set of challenges that are particularly prominent, in healthcare settings. These challenges include demands, the requirement for extensive and well labelled datasets and the complexities involved in deploying such models. Initiatives such as the ISLES challenge have played a role in establishing datasets for research purposes, however there is still a crucial need for more diverse and larger datasets to effectively train models capable of generalizing across different populations. Moreover developing frameworks that're both efficient in terms of computation and adaptable [13], to different clinical scenarios remains an ongoing area of exploration and study.

In the studies could look into enhancing structures to smoothly integrate with current clinical setups in places, with limited resources.It is important to combine IoMT frameworks with models for diagnostic processes through quick data processing and decision making support.Improving the interpretability of models is crucial for giving clinicians insights that build confidence in AI assisted diagnostics.Eventually creating scalable systems will be key, to unlocking the full potential of deep learning and artificial intelligence in improving stroke diagnosis and treatment.

Research methodology

In this research project, a collection of 2,501 and white pictures (650 pixels by 650 pixels) sourced from the available Kaggle stroke dataset was utilized. The collection was split into three parts: 70 %, for training purposes 20 % for validation tasks and 10 %, for testing procedures. The information was classified into two groups. “normal” and “stroke” creating a classification challenge.Training the model to achieve performance was challenging due, to the dataset size which required the implementation of data augmentation techniques to expand the training data artificially and enhance the models overall performance.

To enhance the variety within the training set we implemented data augmentation methods. These techniques comprised rotations, flips along the axis adjustments, in width and height as well, as zooming effects. The selection of these augmentations aimed to encompass viewpoints and boost the models resilience by mimicking real world scenarios. Each modification was randomly introduced during training to ensure that each epoch involved modified renditions of the data, which helped prevent overfitting and promoted generalization.

The main model used in this research was a model that integrated forecasts, from three structures namely CNN, EfficientNetB7 and DenseNet201. The blending method was crafted to utilize the features of each structure possibly offsetting any shortcomings. The structural design of the model as shown in Figure 1 consists of the elements:

1. A unique CNN design featuring convolution and pooling layers tailored for this dataset was developed by CNN.
2. EfficientNetB7 is a performing model that has been trained on ImageNet and is renowned for its adjustments, in depth width and resolution.
3. DenseNet201 enhances the flow of gradients, throughout the models layers and is especially advantageous, for handling datasets.

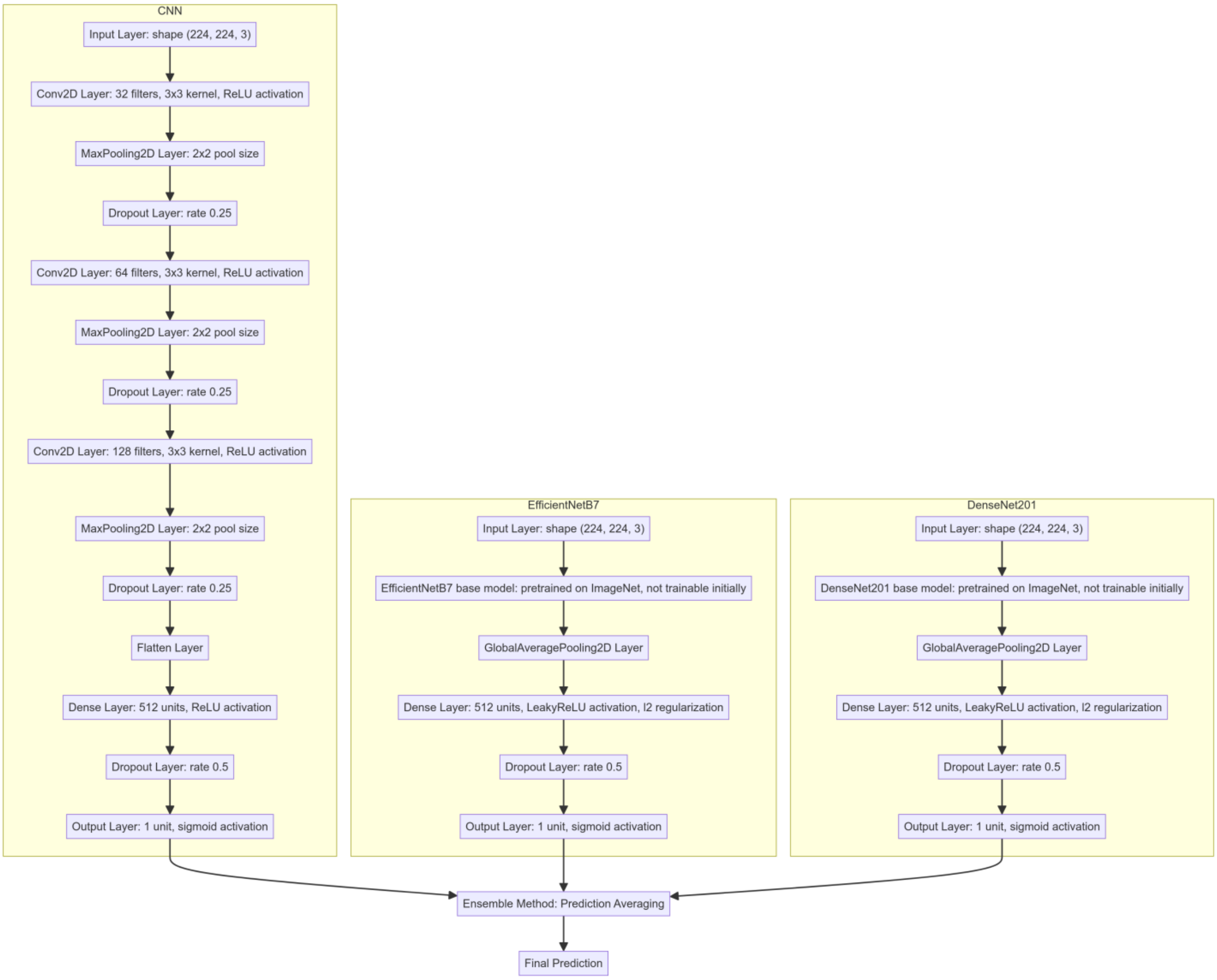

Figure 1. The design of a collective model structure, for classifying brain images

Each model was adjusted to work with the dataset by adding connected layers designed for identifying stroke images accurately. The results, from these models were averaged to create the outcome of the group model that utilizes the advantages of each network.

To assess how well the ensemble model performed in comparison, to the baseline models training and testing processes involved working with four models:

- Mobile Net V2 and Efficient Net B0 used the pre-existing ImageNet. Layers have been added to adapt them for stroke classification purposes. The models were compiled using the Adam optimizer with a binary cross-entropy loss function and accuracy as an indicator of performance measurement.

To optimize the effectiveness of training and prevent problems with retraining during the model training period, adjustments to the learning rate were included in 50 epochs. The mechanisms of early stopping with the help of callbacks are implemented.

- Both ResNet50 and the custom CNN model utilized trained weights that underwent fine tuning by adding extra layers like the other models did too Class weights were also tweaked to handle the imbalance, in dataset classes Both models underwent an initial 10 epochs of training followed by an additional 20 epochs, to fine tune the last layers.

In order to work well with trained models and their weights that are set for a certain size, like 224 by 224 pixels all the images in the research were adjusted to this size for uniform comparison and integration, into the model ensemble.

Results of the study

Here are the main results, from comparing the combined model with structures, on the training set and testing set using accuracy and loss measurements to assess each models performance. MobileNetV2 exhibited the test accuracy of 65 625% as depicted in figure 2 and demonstrated via training and validation loss graphs souvenir whereas EfficientNetB0 also delivered results with a final accuracy score of 59 375%. These outcomes underscore the challenges faced by models when dealing with data sets.

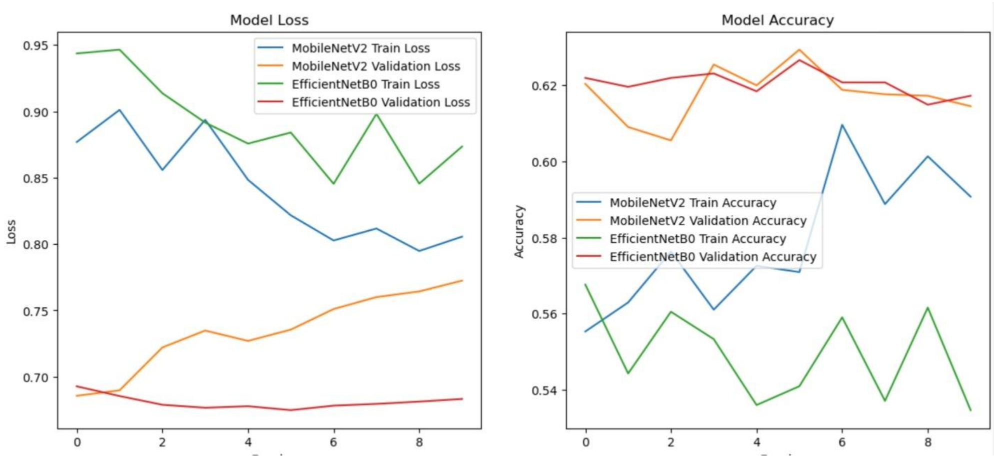

Figure 2. Loss in training and validation of MobileNetV2 and EfficientNetB0 models

After some additional adjustments and fine tuning, to the ResNet50 model in our study experiment showed outcomes with a peak test accuracy of 61,5% The training process details are represented in figure 3 with insights on the variations in both training and validation losses highlighted This evidence suggests that advanced models, like ResNet50 encounter challenges when working with limited datasets.

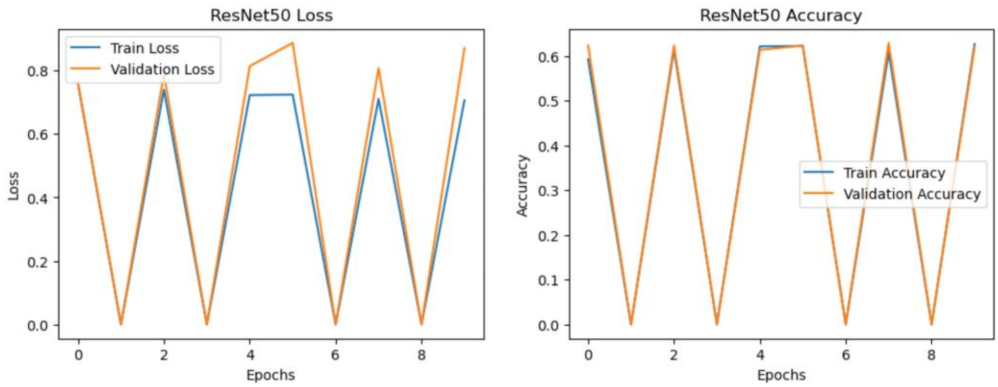

Figure 3. The ResNet50 model showed both train and validation losses during the training process

In an attempt to improve accuracy rates than 61,5% a combination of CNN models, like EfficientNetB7 and DenseNet201 was utilized in an approach as supported by Figure 4. Though ensemble learning is theoretically advantageous in boosting accuracy levels of models in practice; this combination failed to achieve an improvement beyond what was achieved by individual models, like MobileNetV2.

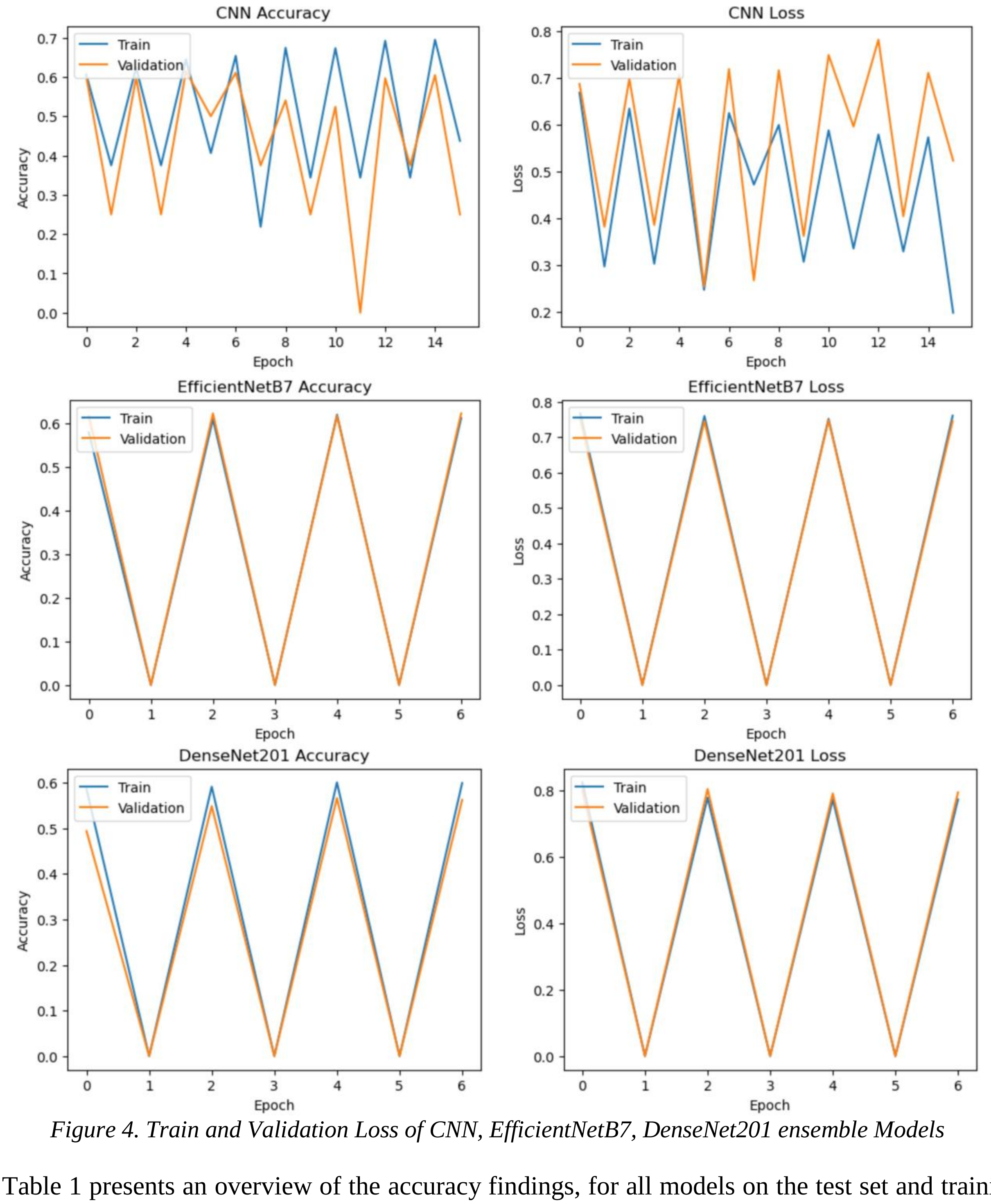

Figure 4. Train and Validation Loss of CNN, EfficientNetB7, DenseNet201 ensemble Models

Table 1 presents an overview of the accuracy findings, for all models on the test set and training sets as the validation dataset results and performance metrics displayed a close match between the top performing individual models and the ensemble models highest test accuracy level which indicates a requirement, for a more extensive dataset to enhance meaningful progress.

Table 1. Results, for accuracy, on the validation and training set for all models.

Model	*Test acc*	*Validation acc*	*Train acc*
MobileNetV2	*65.625%*	*61.83%*	*65.39%*
EfficientNetB0	*59.375%*	*62.07%*	*62.21%*
ResNet50	*61.5%*	*60.48%*	*67.95%*
EfficientNetB7	*62.5%*	*60.48%*	*67.95%*
DenseNet201	*60%*	*60.56%*	*61.72%*
CNN	*61.17%*	*61.91%*	*61.16%*
Ensemble model	*61.5%*	*56.25%*	*80.73%*

Discussion

The results of the experiment suggest that combining models does not always result in enhancements, in accuracy when categorizing data sets. The combined model displayed an enhancement compared to individual models, however these outcomes are constrained by the quantity and variety of the data used. These discoveries align with research that highlights how the effectiveness of combined models is influenced by both the abundance and caliber of the data.

The models potential is greatly hampered by the dataset size. Even though using data augmentation techniques helped broaden the variety in the training set it was not sufficient to make up for the shortage of data. This underscores the importance of expanding the dataset whether through growing the volume of data or utilizing data and other augmentation strategies. In the studies could concentrate on enhancing data augmentation methods and utilizing data to address the challenges of working with a limited dataset size effectively. It might prove beneficial to delve into sophisticated ensemble strategies, like weighted averaging and stacking in conjunction, with other deep learning approaches.

The outcomes coincide with discoveries, in the realm of stroke categorization that underscore the significance of varied data sets in attaining superior precision levels. Moreover numerous research studies have highlighted that sophisticated ensemble approaches yield outcomes across various data sets validating the need for continued enhancements, in this domain.

Conclusion

This research delved into how combining CNN with EfficientNetB7 and DenseNet201 could be used to classify stroke images from a dataset. Even though using architectures, in a model showed promise the results highlighted that the accuracy and reliability of the model were limited by the datasets size and complexity. The test accuracy reached was similar, to what individual baseline models achieved emphasizing the importance of improving both data availability and model refinement in studies.

In order to overcome these constraints in the work of us will concentrate on implementing the following enhancements:

- Data Expansion Strategy involves enlargening the dataset to offer a range of training examples, for the models to learn from and improve their ability to generalize effectively.
- Advanced data augmentation involves using techniques to generate a range of training samples that can assist models in adjusting to the intricate patterns found in stroke images.
- Enhanced Refinement Technique; exploring the inclusion of layers, in existing models to grasp more intricate nuances that might enhance the precision of categorization.
- Utilizing a range of model designs in the ensemble to capture aspects of the data and potentially strengthen the ensembles effectiveness, through diversity, in model integration.
- Fine tuning Parameters Systematically. Adjusting factors like learning rates and batch sizes along with tweaking layer setups to enhance the models effectiveness.

In summary although the ensemble method shows potential, for categorizing strokes achieving strong and practical performance relies on using datasets and improving model settings. Subsequent research in this area focuses on boosting the models precision and dependability which will help create instruments, for managing strokes.

Acknowledgment

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. AP22686812).

References

[1] Bathla P., Kumar R. A hybrid system to predict brain stroke using a combined feature selection and classifier //Intelligent Medicine. – 2024. – Т. 4. – №. 02. – pp. 75-82. https://doi.org/10.1016/j.imed.2023.06.002

[2] Chaki J., Woźniak M. Deep learning and artificial intelligence in action (2019-2023): A review on brain stroke detection, diagnosis, and intelligent post-stroke rehabilitation management //IEEE Access. – 2024. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10485414

[3] Abulfaraj A. W., Dutta A. K., Sait A. R. W. Ensemble Learning-based Brain Stroke Prediction Model Using Magnetic Resonance Imaging //Journal of Disability Research. – 2024. – Т. 3. – №. 5. – pp. 1-9. https://www.scienceopen.com/hosted-document?doi=10.57197/JDR-2024-0061

[4] Feigin V. L. et al. World Stroke Organization (WSO): global stroke fact sheet 2022 //International Journal of Stroke. – 2022. – Т. 17. – №. 1. – pp. 18-29. doi: 10.1177/17474930221080343

[5] Tursynova A. et al. Deep learning-enabled brain stroke classification on computed tomography images //Comput. Mater. Contin. – 2023. – Т. 75. – №. 1. – pp. 1431-1446. https://doi.org/10.32604/cmc.2023.034400

[6] Thiyagarajan S. K., Murugan K. A systematic review on techniques adapted for segmentation and classification of ischemic stroke lesions from brain MR images //Wireless Personal Communications. – 2021. – Т. 118. – №. 2. – pp. 1225-1244. https://doi.org/10.1007/s11277-021-08069-z

[7] Douaud G. et al. SARS-CoV-2 is associated with changes in brain structure in UK Biobank //Nature. – 2022. – Т. 604. – №. 7907. – pp. 697-707. https://doi.org/10.1038/s41586-022-04569-5

[8] Ke Q. et al. Adaptive independent subspace analysis of brain magnetic resonance imaging data //Ieee Access. – 2019. – Т. 7. – pp. 12252-12261. https://doi.org/10.1109/ACCESS.2019.2893496

[9] You W., Henneberg R., Henneberg M. Healthcare services relaxing natural selection may contribute to increase of dementia incidence //Scientific Reports. – 2022. – Т. 12. – №. 1. https://doi.org/10.1038/s41598-022-12678-4

[10] Fan L., Zhang L. Multi-system fusion based on deep neural network and cloud edge computing and its application in intelligent manufacturing //Neural Computing and Applications. – 2022. – Т. 34. – №. 5. – pp. 3411-3420. https://doi.org/10.1007/s00521-021-05735-y

[11] Lo C. M., Hung P. H., Hsieh K. L. C. Computer-aided detection of hyperacute stroke based on relative radiomic patterns in computed tomography //Applied Sciences. – 2019. – Т. 9. – №. 8. https://doi.org/10.3390/app9081668

[12] Shinohara Y. et al. Development of a deep learning model to identify hyperdense MCA sign in patients with acute ischemic stroke //Japanese journal of radiology. – 2020. – Т. 38. – pp. 112-117. https://link.springer.com/article/10.1007/s11604-019-00894-4

[13] Shin H. et al. Automated segmentation of chronic stroke lesion using efficient U-Net architecture //Biocybernetics and Biomedical Engineering. – 2022. – Т. 42. – №. 1. – pp. 285-294. https://doi.org/10.1016/j.bbe.2022.01.002