Deep Learning-Based Classification of Brain Strokes Using Head CT Images: Fine-Tuning and Explainable Artificial Intelligence Analysis of a ResNet18 Model

Tom André [1] Lianne Dame Blok [1] Sem de Groot [1], Stijn Pleunes [1] Woobie Woobiya Reeh [1]

[1] Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

Abstract

Stroke is a leading cause of morbidity and mortality worldwide. Accurate differentiation between ischemic and hemorrhagic (bleeding) stroke on non-contrast computed tomography (CT) is essential, as treatment strategies differ fundamentally. This distinction remains challenging in acute settings, where time is limited. Retrospective studies have reported diagnostic error rates in neuroradiology ranging from 1.7% to 7.7% per case, highlighting the potential of automated tools to support radiologists and reduce errors.

This study evaluated the classification performance of Machine Learning (ML) and Deep Learning (DL) models for detecting stroke types in CT images, categorized as normal, hemorrhagic, and ischemic. The models assessed were a Support Vector Machine (SVM) with Principal Component Analysis (PCA), a basic Convolutional Neural Network (CNN), and a fine-tuned ResNet18 architecture. Performance was measured using accuracy, precision, recall, and F1-score, with F1-score being the most clinically relevant metric. This is because false negatives may result in missed diagnoses, while false positives can lead to incorrect therapy, which is particularly harmful given the opposing treatment strategies required for ischemic and hemorrhagic strokes.

The fine-tuned ResNet18 model achieved the highest performance. Specifically, it reached a test accuracy of 97% and class-specific F1-scores of 0.98 for normal, 0.95 for bleeding, and 0.94 for ischemia. The macro-average F1-score was 0.96, reflecting strong and consistent classification performance across all categories. To further examine model behavior, misclassified cases were analyzed using Explainable Artificial Intelligence (XAI). These errors were often associated with a broader and less specific focus of the model on relevant image regions.

Concluding, The ResNet18 model outperformed the other approaches and demonstrated classification accuracy comparable to expert neuroradiologists, while providing faster inference. These findings support its potential as a clinical decision support tool for acute stroke diagnosis.

Keywords: Stroke CT Classification, Deep Learning, ResNet18, Explainable Artificial Intelligence

Introduction

Stroke is a leading cause of death and long-term disability worldwide. According to the World Stroke Organization (WSO), approximately 12 million people experience a stroke each year, with 6.5 million resulting in fatalities. Notably, over 60% of these strokes occur in individuals under the age of 70, and 16% affect those under the age of 50, indicating the widespread impact across age groups [1]. The burden of stroke extends beyond mortality, as many survivors face significant neurological impairments, placing a substantial strain on patients and healthcare systems.

Strokes are broadly categorized into two primary types: ischemic and hemorrhagic. Ischemic strokes, which account for approximately 87% of all cases, occur due to an obstruction in blood flow to the brain, often caused by a thrombus or embolus [2]. In contrast, hemorrhagic strokes result from bleeding within the brain tissue, typically due to a ruptured blood vessel. The distinction between these types is critical, as their treatments differ significantly. Ischemic strokes may be treated with thrombolytic agents to dissolve clots, whereas hemorrhagic strokes often require surgical intervention to control bleeding. Misclassification may result in inappropriate treatment, potentially worsening patient outcomes. In severe cases, it can even be fatal. For example, administering blood thinners to a patient with a hemorrhagic stroke can cause severe bleeding and significantly compromise the patient's condition [3].

Therefore, accurate and rapid diagnosis is critical in stroke management. Non-contrast computed tomography (CT) scans are the standard initial imaging method due to their speed and accessibility. However, interpreting these scans requires specialized expertise, and subtle differences between stroke types can be challenging to detect, leading to potential delays in treatment. In addition, A retrospective study has reported diagnostic error rates in neuroradiology ranging from 1.7% to 7.7% per study, highlighting the need for automated tools to support radiologists and reduce the risk of misdiagnosis [4].

Recent advancements in Artificial Intelligence (AI), particularly in the field of deep learning (DL) and machine learning (ML), have opened new possibilities for improving diagnostic accuracy in medical imaging. Traditional ML methods such as Support Vector Machines (SVM), especially when combined with dimensionality reduction techniques like Principal Component Analysis (PCA), have shown promise in handling high-dimensional image data efficiently [5]. However, due to the complexity and variability present in medical images, more sophisticated models are often required. Convolutional Neural Networks (CNN), a class of DL models designed specifically for image data, have demonstrated superior performance in classification tasks by learning spatial features directly from the input [6]. ResNet18 is a pre-trained CNN with 18 layers and has shown strong performance in medical image analysis tasks. Its architecture is designed to support effective training of deep models, making it a reliable choice for complex image recognition problems [7].

The aim of this project is to develop an automated tool to classify non-contrast brain CT scans into three categories: ischemic stroke, hemorrhagic stroke, and normal findings, with the potential to assist clinicians in the diagnostic process. Both ML and DL approaches were investigated. First, an SVM model was developed in combination with PCA to manage the high dimensionality of the image data. This was followed by a basic CNN, and later the fine-tuning of a ResNet18 model. All models were trained using preprocessed CT data. The SVM model was optimized through hyperparameter tuning, while the CNN and ResNet18 models were trained using adaptive optimizers and learning rate scheduling based on validation performance. Training progress was monitored throughout to ensure stable convergence and generalization.

To evaluate model performance, accuracy, precision, recall, and F1-score were calculated. The F1-score is considered especially relevant in the context of stroke classification, since false negatives may result in missed diagnoses and false positives can lead to inappropriate treatment decisions. This distinction is particularly critical when differentiating between ischemic and hemorrhagic strokes, as they require opposing treatment strategies. Finally, Explainable Artificial Intelligence (XAI) techniques were employed to gain further insight into model behavior. For this purpose, Gradient-weighted Class Activation Mapping (Grad-CAM) was used to highlight the regions of CT scans that most influenced the model's predictions [8, 9]. These visual explanations help to identify possible sources of misclassification and improve the interpretability of the model, which is essential for model improvement and clinical integration.

Material and Methods

Data Acquisition

For this project, the "Brain Stroke CT Dataset" from Kaggle was used [10, 11]. This dataset contains non-contrast CT scans categorized into three classes: normal, ischemic stroke, and hemorrhagic stroke. It comprises a total of 6650 labeled images, including 4427 normal scans, 1130 scans showing ischemic stroke, and 1093 scans with hemorrhagic stroke. Although the dataset is imbalanced, it is well suited to the objectives of this study due to the inclusion of all relevant stroke types and the availability of clearly labeled data. The images were annotated by radiologists, ensuring clinically meaningful ground truth for supervised model development.

Software

This project was implemented in Python (version 3.12.7), making use of a range of libraries for data processing, model development, evaluation, and visualization. The core deep learning framework used was PyTorch (version 2.5.1), complemented by TorchVision (version 0.20.1) for image preprocessing and augmentation, and TorchCam (version 0.3.1) for generating class activation maps. Scikit-learn was employed for data splitting, model evaluation, and calculation of performance metrics, with support from utilities such as joblib for model persistence and tqdm for progress tracking. Image handling was managed using Pillow, while data manipulation was performed using NumPy and Pandas. For model interpretability, Grad-CAM visualizations were created to highlight relevant regions of the input images. Results and diagnostic plots were generated using Matplotlib (version 3.10.1) and Seaborn (version 0.13.2). Additional dependencies, including python-dateutil, threadpoolctl, and packaging, supported various internal operations of the tools used.

Data Preprocessing

The dataset consists of non-contrast CT scans classified into three categories, with each class stored in a separate folder and assigned a distinct numeric label. No merging of datasets was required, and no missing data or labels were encountered during the loading process.

To prepare the data for the SVM model, images were resized to 64 by 64 pixels to maintain computational efficiency and converted to grayscale to reduce input complexity while retaining relevant structural information. For the deep learning models, larger image dimensions were used: 256 by 256 pixels for the basic CNN, and 128 by 128 pixels for ResNet18.

In the SVM pipeline, image pixel values were flattened and dimensionality was reduced using PCA. For the CNN-based models, relevant features were learned automatically through the training process. Manual feature engineering or selection was not applied, as it is not applicable to the chosen modelling approaches.

Model Architecture

Three types of models were implemented in this study.

First, a SVM was used as a baseline model. Although SVMs are not specifically designed for image data, they can still achieve reasonable performance when trained on flattened pixel inputs. After hyperparameter tuning, the addition of PCA was found to significantly improve classification performance, and was therefore included as part of the SVM pipeline [12].

Second, a basic CNN was developed as an initial deep learning baseline. This architecture consisted of three convolutional layers and served as a comparison point for evaluating the performance of more advanced models [13].

Third, ResNet18 was implemented. This is a pre-trained CNN with 18 layers, based on residual learning. The core component of ResNet18 is the BasicBlock, which includes convolutional layers, batch normalization, and ReLU activation. The model also includes an initial set of layers to process input images, followed by four main residual layers that progressively increase the number of channels and perform downsampling, enabling deeper hierarchical feature extraction [14].

Training and Hyperparameter Optimization

To improve model performance, hyperparameter optimization and regularization techniques were applied across all models.

For the baseline SVM, several preprocessing steps and hyperparameters were tuned. StandardScaler was used to normalize all input features to a common scale, and PCA was applied to reduce dimensionality while preserving the most informative components. The SVM classifier (SVC) was configured to handle class imbalance using balanced class weights. A GridSearchCV procedure with 3-fold cross-validation was employed to explore all combinations of selected hyperparameters, and the best configuration was chosen based on validation accuracy. For each combination, the model was trained and validated three times on different folds of the training set.

- The number of PCA components (*n_components*) was varied across **100**, **150**, and **200**, with higher values resulting in more features for the SVM model.
- The regularization parameter (C) was tested at values of **1**, **10**, and **100**, where higher values correspond to a tighter fit to the training data.
- The kernel coefficient (γ) was evaluated using values **'scale'**, **0.01**, and **0.001**; lower values emphasize broader trends in the data, while higher values allow the model to focus more on specific instances.

For the basic CNN, the model was trained using the Adam optimizer with a learning rate of $1e^{-4}$ and CrossEntropyLoss as the loss function. A ReduceLROnPlateau scheduler was used to decrease the learning rate when the validation loss plateaued. Early stopping was implemented with a patience of 5 epochs, based on validation loss, to prevent overfitting and ensure efficient training.

For the ResNet18 model, additional optimization strategies were introduced. Data augmentation techniques such as random horizontal flips, rotations, and color jittering were applied during training to improve generalization. Class weighting was incorporated into the loss function to address class imbalance and ensure that underrepresented categories received sufficient attention. Learning rate scheduling was again applied to reduce the learning rate when validation performance no longer improved. Early stopping was configured with a patience of 10 epochs, this time based on validation accuracy, as it proved to be a more reliable indicator of generalization to the test set. Compared to the CNN baseline, ResNet18 was trained for more epochs with a higher patience threshold, allowing the model additional time to converge and reach optimal performance.

Evaluation Metrics

Accuracy, precision, recall, and F1-score were used to evaluate model performance. The F1-score was considered most important, as it balances precision and recall and captures the clinical relevance of false positives and false negatives, as discussed in the *Introduction.*

Results

Data Exploration

The class distribution and a subset of CT images from each category were visualized to better understand the dataset and evaluate class balance. *Figure 1* shows a clear class imbalance, with the majority of images labeled as Normal, while the Bleeding and Ischemia classes contain significantly fewer samples. This imbalance may affect model performance by favouring the majority class during training. To reduce this risk, stratified sampling will be applied.

Additionally, *Figure 2* presents example CT images for each class. The images illustrate typical examples of Normal, Bleeding, and Ischemia scans included in the dataset. While detailed pathological features may not always be visually distinguishable, the images provide an overview of the data diversity and confirm the presence of different stroke categories.

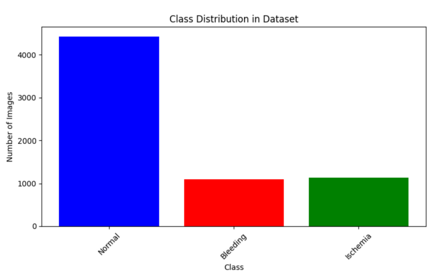

Figure 1. Class distribution of the dataset, showing the number of non-contrast computed tomography (CT) images for Normal, Bleeding, and Ischemia categories.

Support Vector Machine with Principal Component Analysis

To evaluate the performance of the SVM combined with PCA, a confusion matrix and classification report were generated on the test set. The best hyperparameters, identified through grid search, included 150 principal components, an SVM penalty parameter (C) of 100, and a kernel coefficient (γ) of 0.001 (*Table 1*).

Figure 3 shows the confusion matrix for the test set. Most Normal samples were correctly classified (872 out of 886), with only minor misclassifications. The Bleeding class had 179 correct predictions out of 218, while the Ischemia class showed 189 correct predictions out of 226. Misclassifications were relatively few and mostly occurred between Bleeding and Normal, or Ischemia and Normal.

Table 2 summarizes the precision, recall, and F1-scores. The model achieved an overall accuracy of 93%. Precision and recall were highest for the Normal class (0.92 and 0.98, respectively). The Bleeding and Ischemia classes also showed strong performance, each with an F1-score

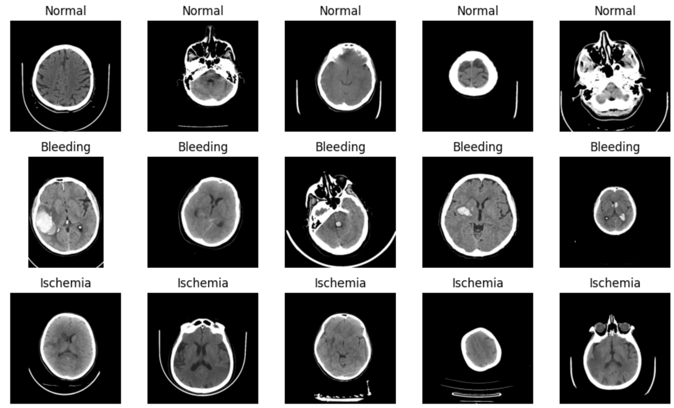

Figure 2. Example non-contrast computed tomography (CT) images from the dataset representing each class: Normal, Bleeding, and Ischemia.

of 0.89. The macro-average F1-score was 0.91, indicating consistent classification performance across classes.

Table 1. Best hyperparameters found for the Support Vector Machine (SVM) with Principal Component Analysis (PCA) model using grid search.

Hyperparameter	Value
Number of PCA components (*n_components*)	150
SVM penalty parameter (C)	100
SVM kernel coefficient (γ)	0.001

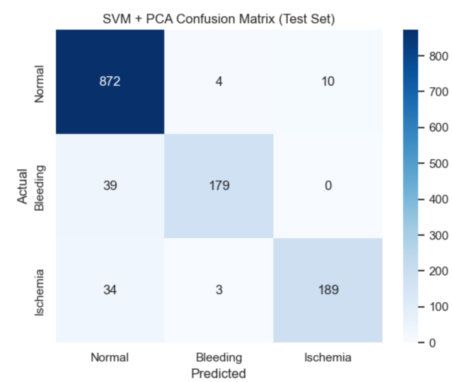

Figure 3. Confusion matrix for the Support Vector Machine (SVM) with Principal Component Analysis (PCA) model on the test set

Table 2. Classification report for the Support Vector Machine (SVM) with Principal Component Analysis (PCA) model on the test set.

Class	Precision	Recall	F1-score	Support
Normal	0.92	0.98	0.95	886
Bleeding	0.96	0.82	0.89	218
Ischemia	0.95	0.84	0.89	226
Accuracy				0.93 (1330)
Macro avg	0.94	0.88	0.91	1330
Weighted avg	0.93	0.93	0.93	1330

Convolutional Neural Network

The training performance of the CNN model was evaluated by monitoring the training and validation loss curves over 20 epochs, with early stopping triggered after 15 epochs (*Figure 4*). Training accuracy steadily increased, nearing 100% by the final epoch, while validation accuracy plateaued around 90%. Although training loss continuously decreased, validation loss stabilized and showed fluctuations after approximately epoch 10, indicating that the model started overfitting on the training data. Overall, convergence was good, but generalization to unseen data became limited beyond epoch 10.

Final classification performance on the test set was evaluated using a confusion matrix and classification report (*Figure 5* and *Table 3*). The model achieved an overall accuracy of 88%. The Normal class performed best, with a precision of 0.91, recall of 0.94, and F1-score of 0.93. The Bleeding class showed a precision of 0.85 and a recall of 0.70, while the Ischemia class reached a precision of 0.77 and a recall of 0.80. The macro-average F1-score was 0.83, indicating reasonably balanced performance across classes.

The confusion matrix shows that most Normal samples (837 out of 886) were correctly classified. Misclassifications occurred more frequently for Bleeding and Ischemia, often confusing these two classes or predicting Normal. These results indicate that the PCA with SVM model outperformed the CNN model in this classification task.

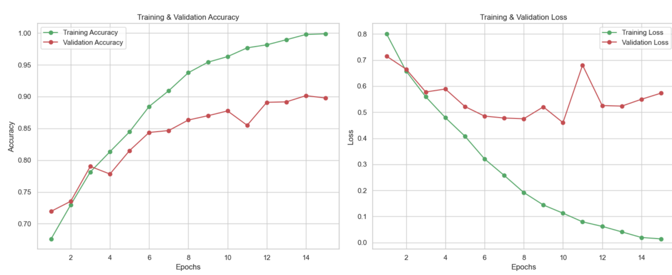

Figure 4. Training and validation accuracy (left) and loss (right) across 15 epochs for the Convolutional Neural Network (CNN) model.

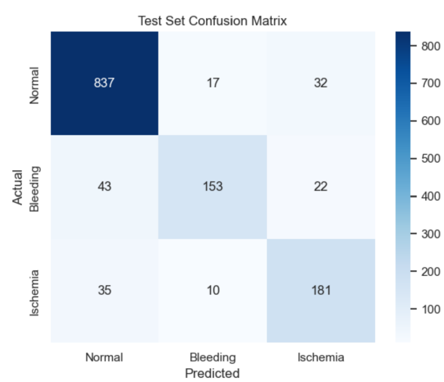

Figure 5. Confusion matrix for the Convolutional Neural Network (CNN) model on the test set.

Table 3. Classification report for the Convolutional Neural Network (CNN) model on the test set.

Class	Precision	Recall	F1-score	Support
Normal	0.91	0.94	0.93	886
Bleeding	0.85	0.70	0.77	218
Ischemia	0.77	0.80	0.79	226
Accuracy				0.88 (1330)
Macro avg	0.84	0.82	0.83	1330
Weighted avg	0.88	0.88	0.88	1330

Fine-Tuned ResNet18 Model

The training and validation performance of the fine-tuned ResNet18 model was monitored across 50 epochs, with early stopping occurring at epoch 44 (*Figure 6*). Training accuracy rapidly reaches 99–100%, while training loss approaches zero, indicating a perfect fit to the training data. Validation accuracy remains consistently high at approximately 96%, while validation loss shows minor fluctuations but remains stable around 0.20–0.25. These results suggest minimal overfitting and overall strong model stability.

Final classification performance was evaluated using a confusion matrix and classification report (*Figure 7* and *Table 4*). The overall accuracy on the test set was 97%. The Normal class achieved a precision of 0.98 and recall of 0.99. The Bleeding class showed a precision of 0.97 and recall of 0.93, while the Ischemia class had a precision of 0.95 and recall of 0.92. The macro-average F1-score was 0.96, reflecting strong and balanced performance across all three classes.

The confusion matrix shows that most samples in each class were correctly classified. Only a small number of misclassifications occurred, mostly between Bleeding and Ischemia. Overall, the fine-tuned ResNet18 model outperformed previous approaches across all relevant classification metrics. To gain insight into the few remaining misclassified cases, an XAI analysis was conducted.

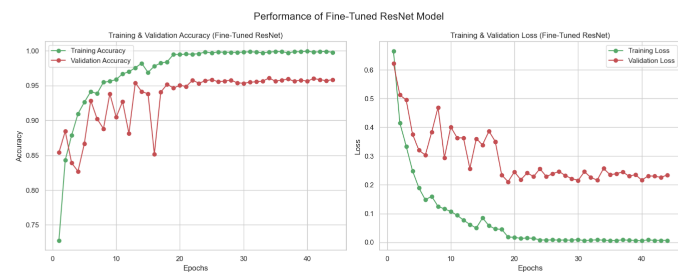

Figure 6. Training and validation accuracy (left) and loss (right) across 45 epochs for the fine-tuned ResNet18 model.

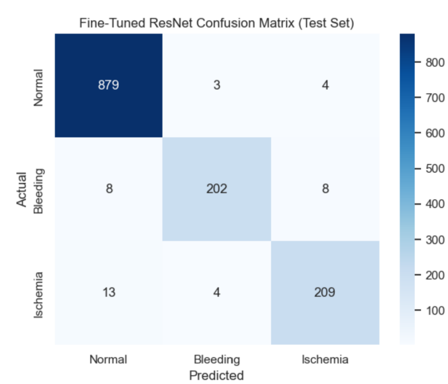

Figure 7. Confusion matrix for the fine-tuned ResNet18 model on the test set.

Explainable Artificial Intelligence of Fine-Tuned ResNet18 Model with Grad-Cam

To better understand how the best-performing model (ResNet18) makes its predictions, Grad-CAM visualizations were created (*Figure 8*). Grad-CAM highlights the areas in the CT images that the model uses to make its decisions. This helps to identify which regions of the brain the model focuses on and to understand why some mistakes may occur.

Table 4. Classification report for the fine-tuned ResNet18 model on the test set.

Class	Precision	Recall	F1-score	Support
Normal	0.98	0.99	0.98	886
Bleeding	0.97	0.93	0.95	218
Ischemia	0.95	0.92	0.94	226
Accuracy				0.97 (1330)
Macro avg	0.96	0.95	0.96	1330
Weighted avg	0.97	0.97	0.97	1330

In the top row of *Figure 8*, the images show correct predictions (true positives). In these examples, the model's attention is directed at specific brain regions, suggesting that it used the correct information to make accurate classifications.

The second and third rows show examples where the model made misclassifications. In these cases, the attention maps are more spread out or focused on less specific areas. This broad focus may have caused the wrong predictions, especially in difficult cases where stroke types can look similar.

Overall, the Grad-CAM results show that the model typically focuses on specific areas when making correct predictions, but its attention becomes less precise when errors occur. These insights help to understand the model's behaviour and may guide future improvements.

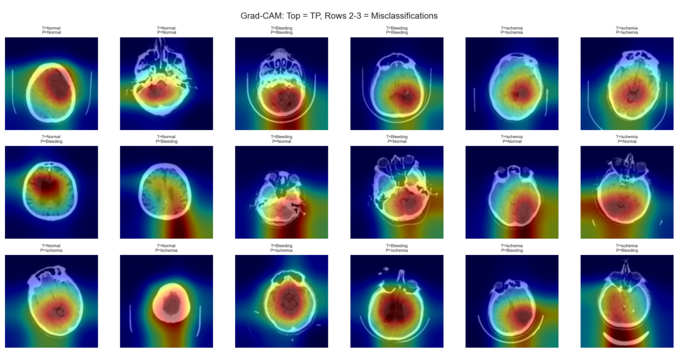

Figure 8. Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations for the fine-tuned ResNet18 model. The top row shows correctly classified images, while rows 2 and 3 show misclassified images.

Discussion and Conclusion

This research investigated the classification performance of both ML and DL models for predicting brain stroke types in CT images, categorized as normal, hemorrhagic (bleeding), or ischemic. Accurate and rapid classification is essential in medical settings to support decision-making and guide treatment, as bleeding and ischemic strokes require different interventions. In addition, a retrospective study has reported diagnostic error rates in neuroradiology ranging from 1.7% to 7.7% per study, underscoring the potential impact of automated tools to assist radiologists and reduce errors [4]. To address this need, a SVM combined with PCA, a basic CNN, and a ResNet18 model were developed and trained. The data was first preprocessed and explored. Hyperparameters were optimized for the SVM with PCA model, while the CNN and ResNet18 models were trained using adaptive optimization techniques and dynamic learning rate scheduling based on validation performance. Besides that, training performance was monitored for the CNN and ResNet18 models. Then, the classification performance of all models was evaluated using precision, recall, F1-score, and accuracy metrics, with F1-score being the most clinically relevant metric. This is because false negatives may result in missed diagnoses, while false positives can lead to incorrect therapy, which is particularly harmful given the opposing treatment strategies required for ischemic and hemorrhagic strokes. Finally, the best-performing model, ResNet18, was further analyzed using XAI with Grad-CAM to better understand how the model makes decisions and how misclassifications occur.

The classification results showed that the fine-tuned ResNet18 model outperformed both the SVM with PCA and the CNN model in terms of accuracy, precision, recall, and F1-score across all stroke categories: Normal, Bleeding, and Ischemia. The ResNet18 model achieved an overall accuracy of 97%, compared to 93% for the SVM with PCA and 88% for the CNN. The ResNet18 also obtained the highest F1-scores, with 0.98 for Normal (compared to 0.95 for SVM with PCA and 0.93 for CNN), 0.95 for Bleeding (compared to 0.89 for SVM with PCA and 0.77 for CNN), and 0.94 for Ischemia (compared to 0.89 for SVM with PCA and 0.79 for CNN). The macro-average F1-score was 0.96 for ResNet18, 0.91 for the SVM with PCA, and 0.83 for the CNN. Training and validation curves further confirmed that the ResNet18 model achieved strong generalization without significant overfitting. These findings align with previous studies where ResNet18 has also outperformed other models in various medical image classification tasks [15]. Grad-CAM visualizations explained these results by showing that the model focused on specific brain regions when making correct predictions, while errors were associated with broader or less specific regions of focus [8, 9].

In conclusion, the ResNet18 model outperformed the other models across all evaluation metrics, achieving particularly strong results on the F1-score. It demonstrated the ability to accurately identify true positives and true negatives across all stroke categories, with performance comparable to that of expert neuroradiologists.

However, this research also has some limitations. First, the dataset exhibited class imbalance, with a higher number of Normal cases compared to Bleeding and Ischemia. Although stratified sampling was applied during data splitting, this imbalance may still have influenced the model's performance. [16]. Second, the models were trained and evaluated on a single dataset, which may limit the generalizability of the results to other populations or imaging settings [17]. Third, although the ResNet18 model demonstrated strong performance, further testing on larger and more diverse datasets is necessary to confirm its robustness. Moreover, limited computational resources restricted the ability to fully optimize the model. Given greater resources, performance could likely be enhanced through extensive hyperparameter tuning (e.g., searching across learning rates, optimizers, and schedulers), model scaling and ensembling (e.g., using ResNet50 or ResNet152), and prolonged training with stronger regularization techniques such as CutMix, MixUp, or label smoothing [18]. Additionally, while Grad-CAM provided valuable insights into the model's decision-making process, it offers only a general visual explanation and may not capture all the factors influencing the model's predictions [9].

Further studies will need to address these limitations to develop an even more robust and accurate model for brain stroke classification in CT images. Still, the fine-tuned ResNet18 model demonstrated strong performance and good generalization, achieving classification accuracy comparable to that of expert neuroradiologists, while offering faster inference. These results highlight its potential as a promising tool to support clinical decision-making in acute stroke diagnosis.

Acknowledgements

The stroke dataset, consisting of cross-sectional CT images, was developed for the Artificial Intelligence in Healthcare competition held in Istanbul in 2021. This competition was organized with support from the Ministry of Health's General Directorate of Health Information Systems and was coordinated by the Turkish Health Institutes (TUSEB) [10, 11].

References

[1] *Impact of Stroke*, https://www.world-stroke.org/world-stroke-day-campaign/about-stroke/impact-of-stroke.

[2] *Types of Stroke and Treatment | American Stroke Association*, https://www.stroke.org/en/about-stroke/types-of-stroke.

[3] J. M. Thon and M. E. Gurol, "Intracranial Hemorrhage Risk in the Era of Antithrombotic Therapies for Ischemic Stroke," *Current Treatment Options in Cardiovascular Medicine*, vol. 18, no. 5, p. 29, May 2016. DOI: 10.1007/s11936-016-0453-y.

[4] G. Biddle, R. Assadsangabi, K. Broadhead, L. Hacein-Bey, and V. Ivanovic, "Diagnostic Errors in Cerebrovascular Pathology: Retrospective Analysis of a Neuroradiology Database at a Large Tertiary Academic Medical Center," *AJNR: American Journal of Neuroradiology*, vol. 43, no. 9, pp. 1271–1278, Sep. 2022. DOI: 10.3174/ajnr.A7596.

[5] J. Li, Z. Weng, H. Xu, Z. Zhang, H. Miao, W. Chen, Z. Liu, X. Zhang, M. Wang, X. Xu, and Q. Ye, "Support Vector Machines (SVM) classification of prostate cancer Gleason score in central gland using multiparametric magnetic resonance images: A cross-validated study," *European Journal of Radiology*, vol. 98, pp. 61–67, Jan. 2018. DOI: 10.1016/j.ejrad.2017.11.001.

[6] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018. DOI: 10.1007/s13244-018-0639-9.

[7] Y.-T. Chen, Y.-L. Chen, Y.-Y. Chen, Y.-T. Huang, H.-F. Wong, J.-L. Yan, and J.-J. Wang, "Deep Learning–Based Brain Computed Tomography Image Classification with Hyperparameter Optimization through Transfer Learning for Stroke," *Diagnostics*, vol. 12, no. 4, p. 807, Apr. 2022. DOI: 10.3390/diagnostics12040807.

[8] V. Kamakshi and N. C. Krishnan, "Explainable Image Classification: The Journey So Far and the Road Ahead," *AI*, vol. 4, no. 3, pp. 620–651, Sep. 2023. DOI: 10.3390/ai4030033.

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020. DOI: 10.1007/s11263-019-01228-7.

[10] *Brain Stroke Dataset*, https://www.kaggle.com/datasets/jillanisofttech/brain-stroke-dataset.

[11] U. Koç, E. Akçapınar Sezer, Y. A. Özkaya, Y. Yarbay, O. Taydaş, V. A. Ayyıldız, H. A. Kızıloğlu, U. Kesimal, İ. Çankaya, M. S. Beşler, E. Karakaş, F. Karademir, N. B. Sebik, M. Bahadır, Ö. Sezer, B. Yeşilyurt, S. Varlı, E. Akdoğan, M. M. Ülgü, and Ş. Birinci, "Artificial Intelligence in Healthcare Competition (TEKNOFEST-2021): Stroke Data Set," *The Eurasian Journal of Medicine*, vol. 54, no. 3, pp. 248–258, Oct. 2022. DOI: 10.5152/eurasianjmed.2022.22096.

[12] *SVC*, https://scikit-learn/stable/modules/generated/sklearn.svm.SVC.html.

[13] *Conv2d — PyTorch 2.7 documentation*, https://docs.pytorch.org/docs/stable/generated/torch.nn.Conv2d.html.

[14] *Resnet18 — Torchvision main documentation*, https://docs.pytorch.org/vision/main/models/generated/torchvision.models.resnet18.html.

[15] W. Xu, Y.-L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107 660, Oct. 2023. DOI: 10.1016/j.cmpb.2023.107660.

[16] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," *Machine Learning*, vol. 113, no. 7, pp. 4845–4901, Jul. 2024. DOI: 10.1007/s10994-022-06268-8.

[17] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep Learning in Medical Image Analysis," *Advances in experimental medicine and biology*, vol. 1213, pp. 3–21, 2020. DOI: 10.1007/978-3-030-33128-3_1.

[18] M. B. Hossain, S. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," *Informatics in Medicine Unlocked*, vol. 30, p. 100 916, 2022. DOI: 10.1016/j.imu.2022.100916.