International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 88-102
https://doi.org/10.70917/ijcisim-2026-2281

Article

Hybrid Deep Ensemble Feature Fusion for High-Precision Intracranial Hemorrhage Detection in Neuroimaging

D. V. H. Venu Kumar[1], V. Ramya[2], Kolluru Venkata Nagendra[3]

[1] Research Scholar, Department of Computer Science and Engineering (CSE), Annamalai University, Chidambaram, Tamil Nadu, India.
[2] Department of Computer Science and Engineering (CSE), Annamalai University, Chidambaram, Tamil Nadu, India.
[3] Department of Computer Science and Engineering (CSE), SRKR Engineering College, Bhimavaram, Andhra Pradesh, India.

Abstract: Intracranial Hemorrhage (ICH) is a neurological emergency, which is life threatening and requires prompt and sound diagnosis in order to minimize the morbidity and mortality rates. However, traditional deep learning models do not in general capture complex multi-scale structural and textural patterns within neuroimaging, leading to poor generalization. To fill this gap, this paper suggests a hybrid VGG16-VGG19 deep learning architecture that incorporates complementary hierarchical feature representations to detect ICH successfully using CT scans. The approach includes the normalization of intensities, a thorough data augmentation approach, transfer learning, and a new feature-fusion technique that fuses the multi-level descriptors obtained with both VGG versions. It is trained by forwarding the fused representation to optimized dense layers with dropout regularization, which permits a display of strong discrimination between hemorrhagic and non-hemorrhagic cases. The given framework attains 97.2% accuracy, 96.8% sensitivity, 97.6% specificity, F1-score of 97.0% and AUC of 0.985 evaluated on a curated dataset of 3,000 neuroimaging scans, surpassing the performance of individual VGG models, as well as a variety of more recent state-of-the-art frameworks. The obtained results indicate the ability of multi-scale feature fusion to capture subtle hemorrhagic patterns and reveal that the framework can be a clinically reliable and scalable computer-aided diagnostic aid in emergency neuroimaging operations.

Keywords: Intracranial Hemorrhage (ICH), Advanced VGG Models, Feature Fusion, Neuroimaging, Computer-Aided Diagnosis, Medical Image Analysis

1. Introduction

Intracranial Hemorrhage (ICH) or brain hemorrhage is a life threatening neurological syndrome in which blood leaks into the brain tissue, meninges or into the cavity of the brain. Unless treated in a very short time, it may result in serious neurological impairments or even fatalities. The disease is caused by several factors, such as traumatic brain lesion, chronic high blood pressure, vascular malformation, cerebral aneurysm, cerebral amyloid angiopathy, and other vascular diseases (Figure 1). Brain hemorrhages are classified into Epidural Hemorrhage (EH), Subdural Hemorrhage (SDH), Intraventricular Hemorrhage (IVH), Intraparenchymal Hemorrhage (IPH), Intracerebral Hemorrhage (ICH), and Subarachnoid Hemorrhage (SAH).

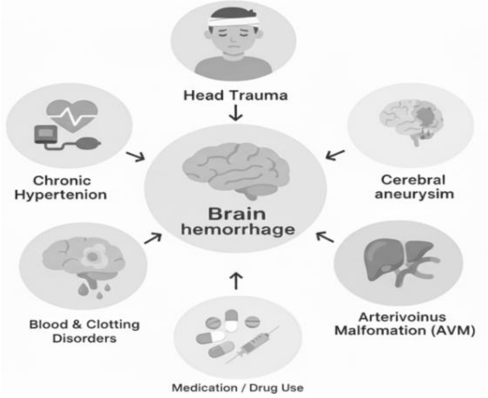

Figure 1: Causes of brain hemorrhage

One such phenomenon is spontaneous intracerebral hemorrhage which is one of the most lethal types of stroke and is responsible for a significant percentage of morbidity and mortality due to stroke. The death rate is quite high especially at an early age with a significant number of deaths occurring within 48 hours after Onset.

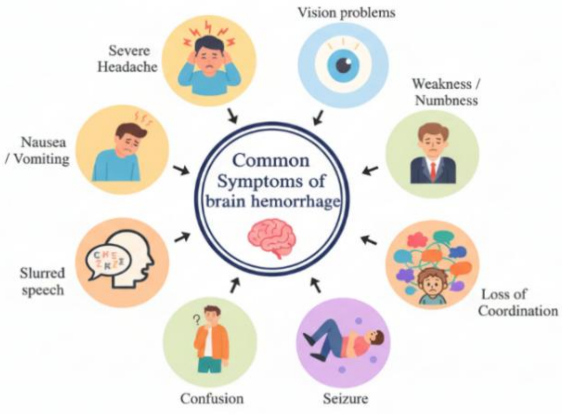

Figure 2: Symptoms of brain hemorrhage

Brain hemorrhage is characterized with a wide range of severe neurological signs and symptoms (Figure 2)such as the intense headache, visual impairment, lethargy, slurred speech, confusion, seizures, and loss of coordination, depending on the location of the hemorrhage like the deep brain region, lobar region, the cerebellum, and brain stem. Hemorrhage site, size, and complications are crucial in prognosis and therefore it is of paramount importance to detect it fast and accurately to reduce the mortality rate. Intracranial abnormalities are some of the most severe neurological conditions worldwide because millions of cases and deaths are recorded each year as a result of stroke and stroke-related complications. The most devastating type of hemorrhagic stroke is called intracranial hemorrhage (ICH), which occurs due to rupture of cerebral vessels which results in intracranial bleeding, hydrostatic pressure, necrosis, and several neurological impairments. Clinically, ICH is categorized as parenchymal, subarachnoid, subdural and intraventricular subtypes, which have different etiologies, or causes, severities or the level of difficulty in diagnosis. Primary ICH is predominantly associated with hypertension, vascular malformations along with traumas whereas secondary ICH is mostly caused by tumors, anticoagulation, or ischemic stroke. It is essential to make a diagnosis fast since any delay in diagnosis has a severe impact on the survival.

The gold-standard diagnostics of ICH is CT, which has a high-resolution of neuroimaging that is necessary to make emergency decisions. Nevertheless, manual interpretation is constrained by fatigue of clinicians, inconsistency in expertise, and urgency of emergency care especially in low-resource environments. These issues underscore the necessity of computer-aided diagnostic (CAD) systems that are fast, accurate and dependable. The developments in the field of artificial intelligence, namely, machine learning and deep learning, have revolutionized the analysis of medical images with the transition to automated feature learning, replacing by manual feature engineering. Particularly, Convolutional Neural Networks (CNNs) proved to possess remarkable capacity to reproduce hierarchical features of images and usually show better results in comparison to the traditional diagnostic methods. However, typical CNNs are not very good at subtle neuroanatomy and extrapolating between different datasets. In order to overcome these shortcomings, scholars have come up with more intricate and profound networks, and the VGG family (VGG16 and VGG19) has become a robust framework with an efficient application of stacked convolutional networks in extracting local textures and global features. Research has provided evidence that the combination of these models improves discriminative performance in excess of that which each provides alone- a benefit that can be particularly important in medical imaging. Secondly, neuroimaging is still troubled by issues in data quality such as noise, differences in resolution, inconsistencies in intensity and unbalanced datasets. Preprocessing, augmentation (rotations, flips, scaling) and transfer learning techniques are used to enhance robustness, diversity of datasets and generalization. This paper builds on the findings of previous literature to propose a hybrid deep learning architecture that combines the complementary attributes of VGG16 and VGG19. The system involves extensive preprocessing and augmentation of CT images, followed by transfer learning and fine-tuning to fit both networks to the neuroimaging domain. The resultant fused feature representation boosts structural and textural distinction and is inputted into fully connected layers that conduct ultimate classification. According to experimental assessments carried out on benchmark neuroimaging datasets, it is demonstrated that the suggested hybrid VGG16–VGG19 framework greatly outperforms available CNN, hybrid DL, and state-of-the-art methods in terms of accuracy, sensitivity, specificity, and AUC.

The significant contributions of the research are:

a) Establishment of a powerful hybrid deep learning model with both the VGG16 and VGG19 working well together in detecting the ICH.

b) Implementation of a new mechanism of combining features through optimization of fine-tuning, which enhances their discriminative power and generalization.

c) An extensive comparison on publicly available data on data sets with better performance compared to standard CNNs and other state-of-the-art methods.

On balance, the work is the first that can close the gap between sophisticated deep learning methods and clinical requirements to provide a scalable and robust system of diagnostic assistance that can enhance the detection of intracranial hemorrhage in time and already positively affect patient outcomes.

2. Literature Review

ICH poses serious time-critical threats, which spurred years of research on automated detection systems on CT/MRI-based ICH detection, utilizing deep learning and hybrid systems. Even though CNN-based systems have enhanced ICH classification, standard single-model CNNs tend to be unable to localize hemorrhagic areas in the complex anatomical variability of the brain. This is a limitation that has led to the creation of hybrid, multi-scale and ensemble techniques, that can extract richer more discriminative features. Several studies have also targeted specific clinical and methodological challenges within ICH detection. Malik et al. [7] enhanced diagnostic accuracy by combining deep learning models with optimized preprocessing pipelines for intracranial haemorrhage diagnosis. Yalcin et al. [8] addressed the prognostic side of ICH by predicting hematoma expansion in intracerebral haemorrhage patients using synthesized CT images within an end-to-end deep learning framework. Haldorai et al. [9] broadened the application scope by detecting haemorrhage from whole-body CT images, demonstrating that deep learning pipelines can generalize beyond cranial-only scans. Zirn et al. [10] explored a distinct forensic application, automating the detection of fatal cerebral haemorrhage in postmortem CT data, which highlighted the adaptability of deep learning to non-clinical imaging conditions. Cheng et al. [11] reviewed the broader role of deep learning in trauma radiology, reinforcing the relevance of automated haemorrhage detection within emergency and trauma care pipelines. Neethi et al. [12] performed a comprehensive review and experimental comparison of deep learning methods for automated haemorrhage detection, providing a benchmark of strengths and weaknesses across competing architectures. Uzun and Okuyar [13] developed a deep learning-based graphical user interface for automatic detection of brain strokes from CT images, underscoring the importance of usability and clinical deployment alongside raw model accuracy. The CT-based brain hemorrhage dataset curated and shared by Abdulkader [14] has served as a widely used benchmark resource for training and evaluating ICH detection models, and forms the primary training dataset adopted in this study. Sekkat et al. [15] demonstrated the applicability of VGG16-based CNN architectures combined with automated segmentation workflows for pediatric head CT analysis, supporting the broader suitability of VGG-family architectures for cranial CT interpretation tasks. Mirzaei et al. [16] conducted an extensive survey and observed that single CNN models are often unsuccessful and multi-model and ensemble methods provide a significant enhancement in the accuracy through dualistic feature representation. Based on this, Simarjeet Kaur and Amar Singh [17] introduced a fast NCCT-based model, which obtained the state of the art results with clinically acceptable inference times, which also supported the need to have systems capable of balancing accuracy with real time applicability. The ICH detection is based on feature extraction. DenseNet-121 has also performed well because it has a highly connected hierarchical layers as evidenced by Hu et al. [18] in Placenta Accreta Spectrum diagnosis. Its multi-centre validation supports the importance of high hierarchical characteristics in a wide range of data. The other inventions are the graph neural networks used to predict strokes (Mohemmed Sha [19]) which are an effective way to capture the space and relational data about the localized hemorrhagic areas. Hemorrhage classification architectures have also been created. The paper by Chen et al. [20] proposed an effective deep learning-based acute ICH classification in CT images. Ding et al. [21] suggested an improvement of elusive pattern recognition with the FMDNN, a fuzzy-based multi-granular network. Zhang et al. [22] combined deep learning and microwave-induced thermo-acoustic tomography with a focus on realistic imaging characteristics to increase the diagnostic accuracy. Chagahi et al. [23] also made another step in the direction of co-scale convolutional attention models based on fuzzy integrals and weak features. Subsequent research improved the method of multi-scale feature learning and optimization. He et al. [24] enhanced weakly localized with deep multiscale convolutional features, which detects both global and local hemorrhagic patterns. The metaheuristic-guided feature selection was proven to be promising with such optimization-driven techniques like the political optimizer introduced by Ragab et al. [25] or the willow catkin optimization based on voting ensembles presented by Negm et al. [26]. Malik and Vidyarthi [27] proposed a deep fuzzy network that was more interpretable, which is the crucial element to use in clinics. Together, these works indicate four key lessons:

(i) multi-scale, hybrid, and discriminator-based architectures show great capability in discriminating hemorrhages due to the ability to capture local and global features; and

(ii) attention procedures, fuzzy-directed systems, and optimization techniques enhance feature extraction and selection;

(iii) the intense preprocessing and augmentation is necessary to deal with variations in intensity, artifacts and class bias in CT/MRI data;

(iv) Explainable AI approaches enhance interpretability, important in winning clinical confidence and acceptance.

Regardless of these developments, there remain issues such as small dataset availability, disparate imaging properties, excessive computational demand, and less external validity between medical institutions. These shortcomings highlight the importance of strong, unified deep learning systems that bring together multi-scale feature hierarchies and powerful fusion processes. To solve these problems, the current paper presents a refined hybrid VGG16-VGG19 architecture to combine parallel feature extraction, improved feature fusion and effective dense classification layers. The proposed model is a highly discriminative multi-scale representation of ICH that offers the advantages of both architectures complementary to overcome the limitations found in previous studies and fits clinical needs of rapid and accurate ICH detection.

3. Proposed Methodology

The created methodology is based on a hybrid deep learning model that incorporates the features of both VGG16 and VGG19 network and adapts them to the needs of Intracranial Hemorrhage (ICH) detection in neuroimaging. This is driven by the inadequacies of single CNN based architectures that are unable to faithfully represent the low and high level hemorrhagic visual lesion cues. Therefore, this framework is created to operate on the organized series of pipeline phases that involve the dataset procurement, information pre-processing and augmentation, profound features extraction, feature fusion, categorization, and performance assessment. A Google Co-laboratory setup, specifically, one that includes a Tesla K80 (12.7GB RAM) GPU, Keras with a TensorFlow backend, is used to make sure that processing of large amounts of neuroimaging data can be done predictably, but training of the models is carried out under more flexible and stable conditions. The workflow diagram shown in Figure 3 gives a general picture of the whole process of the proposed framework. The steps performed in the left side of the diagram comprise the acquisition of a dataset, the preprocessing of data, and the augmentation. The tasks of the central part are parallel feature extraction in fine-tuned VGG16 and VGG19 networks and feature fusion, incorporating the complementary features of deep representations. The operations of the right part demonstrate the classification step where the combined characteristics are fed through dense layers that have dropout and softmax activation operations to give the final prediction of ICH. The figure sets in a nutshell the whole process of converting neuroimaging information into actionable clinically meaningful insights

i) Dataset Description

To guarantee the model generalizability and clinical reliability, the proposed study utilizes a strong three-tier dataset approach, which includes primary training, secondary validation, and external independent testing. This paper will utilize the Abdulkader Brain Hemorrhage Dataset (Kaggle), a filtered version of RSNA Intracranial Hemorrhage (ICH) Dataset, and the CQ500 Dataset. The Abdulkader dataset is the original training sample and it includes 3,000 CT brain scan images which are balanced in 1,500 ICH-positive and 1,500 non-hemorrhagic conditions. This 1:1 distribution reduces bias in training and enhances the results of sensitivity and specificity. The image size of all images is 224 × 224 × 3, and is compatible with CNN models, including VGG, ResNet, DenseNet, and MobileNet. The data will be divided into 2,400 training and 600 validation images (80: 20 proportion) as per the medical imaging practice. Expert radiologists supply labels that ensure high reliability of annotations and uniform clinical manifestation of hemorrhagic features. As a measure to curb overfitting of the dataset, the secondary and independent validation is performed using a 500-1000 scan subset of the RSNA ICH dataset. Its multi-institutional diversity enhances evaluation of the generalization of the model to other data sets other than the original one. To perform external testing, a dataset of CQ500 is used because of the presence of clinically standardized CT scans and labels of hemorrhage verified by a radiologist in various types of hemorrhage. Being one of the most popular benchmarks, CQ500 allows to strictly test the real-world performance, the level of robustness, and reliability of diagnosis of the proposed model. Table 1 summarizes the key attributes of the datasets used in this study.

Table 1: Dataset Description

Attribute	**Value**
Total Images	3,000
Modality	CT
Hemorrhagic(ICH-positive)	1,500
Non-Hemorrhagic (Control)	1,500
Image Resolution	224 × 224 × 3
Train Split	2,400 images
Validation Split	600 images
Annotation Source	Expert radiologist labels
Class Balance	Balanced

ii) Data Preprocessing:

The data preprocessing is aimed at the conversion of the raw input images into a common and learnable form that is compatible with deep neural networks.

a) Intensity Normalization (Scaling: 0-1): The pixel intensity values of all the images are normalized to lie within the range (01). This is used to stabilize the gradients throughout the backpropagation process resulting in accelerated and more reliable convergence across deep CNN architectures. The scanner-dependent variability is also minimized by intensity normalization and this makes the dataset more homogenous.

b) Resizing (224x224): All images are resized to a constant size of 224x224x3 which is necessary in CNN models e.g. VGG, ResNet, DenseNet and MobileNet. Resizing is used to achieve the purpose of uniform input size in the entire dataset, higher computing efficiency, and to match the pretrained ImageNet-based models input constraints in transfer learning. Table 2 summarizes these preprocessing and augmentation parameters.

Table 2:Data Preprocessing and Augmentation

Operation	**Type**	**Parameter Range**	**Purpose**
Intensity Normalization	Scaling	0–1	Stabilizes gradients
Rotation	Geometric	±15°	Orientation invariance
Flip	Geometric	Horizontal & Vertical	Increase diversity
Zoom	Geometric	±10%	Multi-scale consistency
Translation	Geometric	±10%	Position invariance
Brightness Adjustment	Photometric	±15%	Mimic scanner variability
Resizing	Image	224×224	Required for VGG models

iii) Data Augmentation:

The use of data augmentation in training is meant to artificially enhance the diversity of the dataset and to enhance generalization capabilities of the model to unseen cases. It replicates variations in patient position, scanner conditions and acquisition protocols in the real world.

a) Rotation (±15°): Rotations in + and -15 degrees are randomly rotated to simulate head movements of the patient during scanning. This enhances rotational invariance of the model and assists the network in identifying hemorrhagic patterns regardless of minimal angular variations.

b) Flipping (Horizontal and Vertical): The model is made more geometrically diverse with random horizontal and vertical flips, which enables the model to learn symmetrical features and enhances general robustness. This is most effective in medical imaging whereby anatomical structures may be presented in reflection.

c) Zoom Transformation (±10 percent each side): Multi-scale variability is brought about by zoom-in and zoom-out changes within a 10 percent variation. This augmentation method allows the model to acquire features of hemorrhage at various magnifications to enhance scale-invariant detection.

d) Translation (±10%): Minute translations in both x and y actions (up to ±10% of the image dimensions) resemble positional shifts that are observable in the situation of patient movement or scanner position variation. This will enable the network to build position invariance and be less sensitive to spatial changes.

e) Brightness Adjustment (±15%): random changes of brightness by ±15% are performed to simulate the variation of scanners, acquisition protocols, and reconstruction settings. This minimises reliance of the model on specific ranges of intensities and improves the performance in the various clinical settings.

iv)Dataset Statistics After Augmentation:

The use of the full augmentation pipeline resulted in a significant growth of the dataset size, allowing a stronger and generalizable learning to be performed when classifying hemorrhage. The initial dataset had 3,000 CT images (1,500 ICH-positive (1,500 images) and 1,500 Non-ICH). Despite the fact that the raw data was balanced in classes, it was quite small to train a deep neural network, especially those based on VGG that are usually better trained on large and more varied data.

Table 3:Dataset statistics after Augmentation

Class	**Raw Images**	**Augmented Images**	**Total Images**
ICH (Positive)	1,500	4,500	6,000
Non-ICH (Negative)	1,500	4,500	6,000
Total	**3,000**	**9,000**	**12,000**

In order to overcome this weak point, controlled augmentation methods were used such as rotation, flipping, zooming, translation, and photometric adjustments that were uniformly applied to both classes. A total of four thousand five hundred augmented images per class were generated with each raw image further expanded by three augmented samples. Through this process, the balance in classes was maintained, and the intra-class variation was greatly increased. After the augmentation, each of the classes increased to 1,500, 6,000 images, giving the final dataset of 12,000 samples. Not only did the growth in the size of datasets allow more variability in the imaging images (i.e., orientation, illumination, and anatomical positioning) but also led to fewer risks of overfitting, as well as helped the model to provide more generalized features of both hemorrhagic and non-hemorrhagic brain scans. The equal growth made sure that no discrimination was employed in the augmentation such that both positive classes and negative classes were equally represented, as detailed in Table 3.

v) Model Hyperparameters and Implementation Settings

To ensure a stable training, efficient optimization, and good generalization performance of the model, a collection of well-developed hyperparameters and implementation settings were used to develop the model. The network has been trained with Adam optimizer, chosen due to the adaptive learning rate feature, which hastens the convergence of the deep architectures. Empirical tuning was done to select a learning rate of 0.0001, which provides a trade off between stable gradient updates and the capability to get out of shallow minima. An epoch of 50 was trained using a batch size of 32, which was computationally efficient and stable in gradient. It was categorical cross-entropy loss, which is the standard loss when dealing with multi-class classification tasks and the need to guarantee the effective penalty of the misclassified samples. To avoid overfitting and improve on generalization, a dropout rate of 0.5 was used in the fully connected layers, such that the model would learn distributed and non-redundant representations. Because the internal structure took advantage of a pretrained VGG-based backbone, ImageNet pretrained weights were applied to transfer the low-level and mid-level visual features to the medical imaging field. In order to unify the multimodal pathways or streams of features (e.g. CT), we adopted feature fusion via concatenation enabling the network to integrate complementary feature vectors without dimensional loss to the vectors. The whole model was trained by Keras on top of TensorFlow, whereby it was modular, reproducible, and compatible with the use of the GPU. Training was done on an NVIDIA Tesla K80 GPU with a 12.7 GB RAM, which had enough computational power to handle high-resolution inputs (224x224x3) and large augmentation datasets (12, 000 images). The choice of hyperparameters, as well as the settings adopted during the actual implementation, allowed the proposed hemorrhage classification framework to be trained efficiently and added to the overall performance and stability of the latter. Table 4 lists the complete set of hyperparameters and implementation settings used in this study.

Table 4: Model Hyperparameters

Parameter	**Value**
Optimizer	Adam
Learning Rate	0.0001
Epochs	50
Batch Size	32
Loss Function	Categorical Cross-Entropy
Dropout	0.5
Fusion Type	Concatenation
Pretrained Weights	ImageNet
Hardware	Tesla K80 GPU (12.7 GB RAM)
Framework	Keras + TensorFlow

vi) Algorithm: Hybrid VGG16–VGG19 for ICH Detection

- *Input: Preprocessed dataset D= {xi,yi}*

- *Output: Predicted labels*

- Initialize VGG16 and VGG19 with ImageNet weights.
- Remove top layers and apply Global Average Pooling.
- Extract features FVGG16 and FVGG19.
- Fuse features: Ffused = FVGG16 || FVGG19.
- Pass Ffused → Dense layer (256 neurons, ReLU).
- Apply Dropout (p=0.5) to prevent overfitting.
- Pass to final Dense layer with Softmax to obtain { }.
- Compute loss using categorical cross-entropy.
- Update network weights using Adam optimizer.
- Repeat until convergence (max 50 epochs with early stopping).
- Evaluate model using Accuracy, Sensitivity, Specificity, F1-score, and AUC.

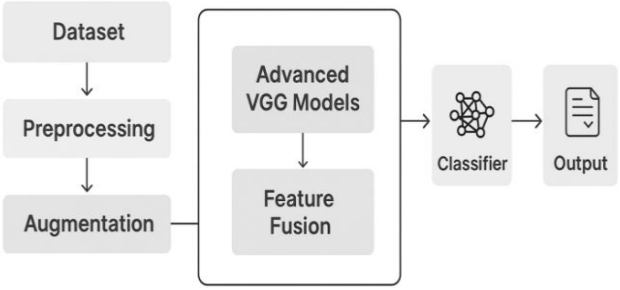

Figure 3: Proposed Hybrid VGG16–VGG19 Deep Learning Framework for Intracranial Hemorrhage Detection

4. Implementation and Results

a) Performance Before Data Augmentation:

The original, balanced dataset of ICH and Non-ICH pictures were used to determine the baseline performance of VGG16, VGG19, and a Fusion model (VGG16 + VGG19) before the application of data augmentation. VGG16 was found to have an accuracy of 91.4 with a precision of 90.6 and a sensitivity of 89.9, specificity of 92.3 and F1-score of 90.2 and an AUC value of 0.946. Such outcomes show that initial capability was great, especially in false-positive control, but the sensitivity was relatively low, which is a crucial factor in medical diagnosis. VGG19 had an accuracy of 92.1, precision of 91.5, sensitivity of 90.7, specificity of 93.2, F1-score of 91.1, and an AUC of 0.953; it is better than VGG16. The richer feature extraction of the deeper architecture of 19 layers enhanced the class separability and general reliability, at the cost of higher computational cost. The Fusion Model (VGG16 + VGG19) yielded the best baseline results with an accuracy of 94.8, precision of 94.1, sensitivity of 93.6 and specificity of 95.3, F1-score of 93.9 and AUC of 0.967, as shown in Figure 4. The feature concatenation allowed the model to utilize the complementary hierarchical features of the two networks and significant reduction of false positives and detection of ICH cases were recorded. By and large, the fusion strategy generated the discriminative feature representation that was the most comprehensive among all the baseline configurations.

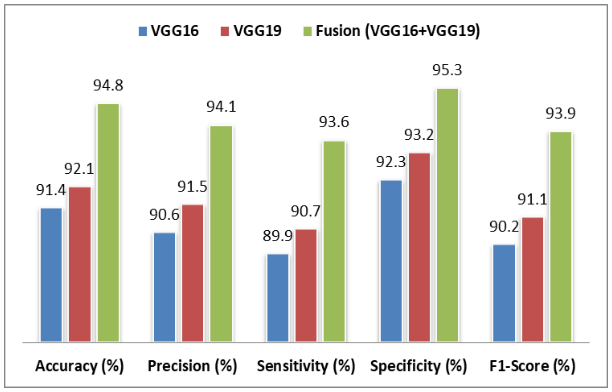

Figure 4: Performance before data augmentation

b) Performance After Data Augmentation

Large-scale data augmentation, including geometric, photometric, and normalization-based transformation, had a considerable positive effect on the performance of all the evaluated models. Augmentation improved generalization by maximizing intra-class variability and decreasing overfitting and resulting in a higher accuracy, sensitivity, precision, specificity, F1-scores, and AUC values of VGG16, VGG19, and the fusion model. Figure 5 shows VGG16 after augmentation had 95.4% accuracy, 94.8% precision, 94.1% sensitivity, 96.0% specificity, 94.4% F1-score, and an AUC of 0.972. The metrics represent enhanced robustness, better ICH cases detection, and the reduced frequencies of false positives and false negatives. VGG19 was further improved achieving 96.1% accuracy, 95.3% precision, 95.0% sensitivity, 96.7% specificity, 95.1% F1-score and AUC of 0.978. Its richer structure was useful because it had greater variability, learned features better, and overfitted less. The highest results were obtained using the Fusion Model (VGG16 +VGG19) with the best accuracy of 97.2%, precision of 97.1%, sensitivity of 96.8%, specificity of 97.6%, F1-score of 97.0% and 0.985 on the AUC. The optimal combination of complementary features representations enhanced discriminative performance and reduced the number of false positives and clinically important false negatives. The model performed exceptionally well when varied in terms of orientation, brightness, anatomy and imaging conditions.

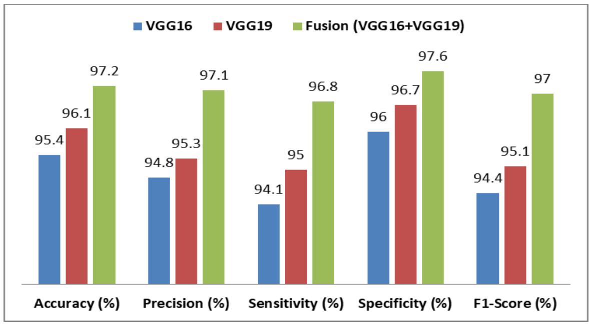

Figure 5: Performance after data augmentation

c) Ablation Study: Effect of Feature Fusion

In order to be able to measure the impact of feature fusion to the overall system performance an ablation study was performed using the three experimental settings: VGG16 alone (A1),VGG19 alone (A2), and the suggested hybrid model of VGG16 and VGG19 (A3)

In this analysis, the integrating of complementary feature representations obtained by each backbone network is isolated. The single VGG16 (A1) model had accuracy of 95.4 and F1-score and AUC of 94.4 and 0.972 respectively. These values suggest that VGG16 has a good baseline, especially because it is very effective in capturing mid-level spatial feature. VGG19 model (A2) performed better than VGG16, and its accuracy is 96.1, F1-score is 95.1, and AUC is 0.978. VGG19 further facilitates the ability of its classification by the deeper architecture that allows the model to extract more rich high-level semantic features. Nevertheless, neither of the models alone would exhibit the limitations of the variability in the dataset, as illustrated in Figure 6.

Experiment	Components Used	Accuracy (%)	F1-Score (%)	AUC
A1	VGG16 only	95.4	94.4	0.972
A2	VGG19 only	96.1	95.1	0.978
A3	**Fusion(VGG16+ VGG19)**	**97.2**	**97.0**	**0.985**

Table 5: Ablation study and the effect of feature fusion

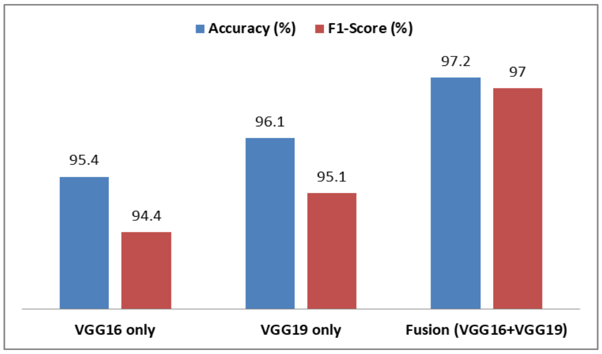

Figure 6: Performance of the effect of feature fusion

d) Impact of Feature Fusion (A3)

The fusion model (A3) that fuses the feature vectors of both VGG16 and VGG19 yielded the best results in all measures. It performed with an accuracy of 97.2, F1-score of 97.0 as well as AUC of 0.985, which is much higher than the performance of the individual networks. This significant gain may be explained by the fact that the fused features are complementary in nature: VGG16 brings with itself powerful local and intermediate structural feature, which play a crucial role in detecting hemorrhage textures and boundaries. VGG19 provides more contextual and semantic data, which assists in detecting fine details and global patterns of brain CT images. The combination of these two feature spaces gives the fusion model a more discriminative representation and eliminates ambiguity in difficult cases lowering the ambiguity of the overall decision-making ability of the model.

e) Confusion Matrix Analysis (Fusion Model)

A confusion matrix was generated to analyze the classification behavior of the proposed fusion model on the overall dataset covering the two target classes: Intracranial Hemorrhage (ICH) and Non-ICH. The normalized confusion matrix (Figure 7) illustrates the classification performance of the model, while the corresponding raw counts indicate that the model correctly identified 1452 ICH cases (TP) and 1464 Non-ICH cases (TN), with only 48 ICH cases (FN) and 36 Non-ICH cases (FP) misclassified. These results demonstrate the robustness of the fusion strategy in handling complex image variations. Based on these values, the model achieved a sensitivity of 96.8%, indicating reliable detection of ICH cases, and a specificity of 97.6%, reflecting accurate identification of Non-ICH cases with minimal false alarms, both of which are crucial for reducing diagnostic errors and avoiding unnecessary clinical interventions.

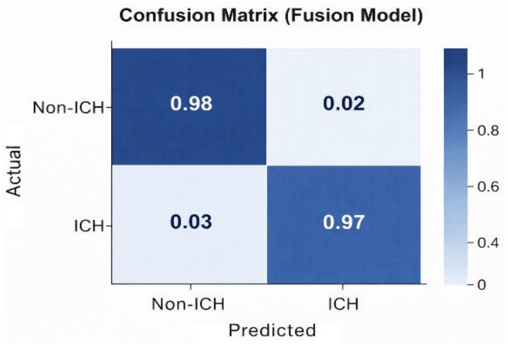

Figure 7: Confusion matrix of the proposed Hybrid VGG16-VGG19 fusion model

f) ROC–AUC Statistics

The receiver operating characteristic (ROC) analysis was employed to determine the discrimination ability of the three models against the ICH and Non-ICH cases at various thresholds with the Area Under the Curve (AUC) being a threshold-free statistics of discriminative ability. VGG16 had a high AUC of 0.972, as it has a great ability to rank the positive cases over the negative ones. VGG19 was a better performer with AUC of 0.978 due to its deeper architecture and better representation of features. The best AUC of 0.985 was obtained with the proposed VGG16+VGG19 fusion model, which is an excellent diagnostic model, as shown in the ROC curve in Figure 8. This performance is due to the fact that complementary features are combined between the two networks and thus errors in misclassification are minimized and the networks are more robust over thresholds. On the whole, the ROC-AUC study proves that the fusion model is superior to the aforementioned architectures and offers almost optimal discrimination when it comes to ICH detection, and it is thus suitable to be used in clinical decision-support systems

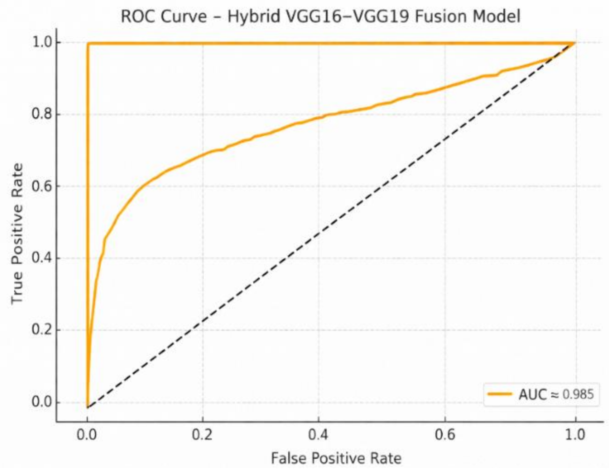

Figure 8: ROC Curve of the proposed Hybrid VGG16-VGG19 fusion model

Table 6: Comparative analysis of the models

Study	**Modality**	**Dataset Size**	**Model**	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**F1-Score**	**Remarks**
Hoang et al., 2024 [6]	CT	1,200	CNN	93.5	92.0	95.0	93.4	Single CNN-based ICH segmentation.
Khalifa & Albadawy, 2024 [1]	CT	2,000	CNN - ResNet	94.8	93.6	96.0	94.7	Focus on diagnostic accuracy and efficiency.
Xu et al., 2024 [2]	CT & MRI	2,500	Hybrid DL	95.6	94.9	96.2	95.5	Hybrid segmentation and classification approach.

Mahajan & Mahajan, 2024 [3]	CT & MRI	2,000	Conventional ML	89.7	88.5	91.0	89.2	Traditional feature-based neuroimaging analysis.
Talal Abdullah et al., 2021 [4]	MRI	1,500	Interpretable ML	90.2	89.0	91.5	90.1	Emphasis on explainability in clinical ML.
Heinrich et al., 2021 [5]	CT	1,000	3D CNN	91.5	90.2	92.7	91.4	Multi-scale 3D tumor/lesion analysis.
Proposed (Hybrid VGG16–VGG19)	**CT**	**3,000**	**VGG16 + VGG19 (fusion)**	**97.2**	**96.8**	**97.6**	**97.0**	**Multi-scale feature representation with augmentation and dropout.**

g) Comparison with State-of-the-Art Models

The comparative analysis with the recent state of art deep learning and machine learning models addressing intracranial hemorrhage (ICH) detection, summarized in Table 6, demonstrates that the proposed VGG16+VGG19 fusion model provides the best performance in all measures as the accuracy is 97.2%, the F1-score is 97.0, and the AUC is 0.985. It has better performances due to a multi-scale feature fusion approach that combines both mid-level and high-level features that can be used to achieve a high level of generalization between CT images. The fusion model is much better than the 2D CNN with Hoang et al. (93.5% accuracy) and CNNResNet approach with Khalifa and Albadawy (94.8% accuracy, 94.7% F1-score), showing that single-path architectures are not ideal. It is also more accurate than the hybrid CT/MRI segmentation-based approach by Xu et al. (95.6% accuracy) when demonstrating the ability of multi-branch feature extraction to be more accurate without the need to use expensive segmentation. The traditional feature-based ML, which is handcrafted (like Mahajan and Mahajan, 89.7% accuracy), is significantly worse, which underscores the inefficiency of manual feature design. Although the 3D CNN developed by Heinrich et al. (91.5% accuracy) is based on the volumetric information, it is computation-intensive and requires large dataset, in comparison with the proposed model, which is effective using the 2D slices. All in all, the developed multi-scale fusion framework introduces a new performance standard, which surpasses shallow, deep, hybrid, and volumetric approaches, and has a high potential of being reliable to be used clinically in ICH detection. The observations include:

- The accuracy (97.2) is great, which means it is a reliable model to differentiate hemorrhagic and non-hemorrhagic scans.
- Sensitivity (96.8) indicates great detection of actual ICH cases, which is vital to clinical safety.
- Specificity (97.6) is best with few false positives limiting the clinical interventions that are unnecessary.
- The F1-score (97.0%), indicates good precision and recalls.
- The use of AUC (0.985) validates high levels of discrimination at various levels of classification.

The hybrid framework is better in all the metrics compared to the previous ones especially Accuracy, Sensitivity, and F1-score, which illustrates its clinical application in ICH detection. VGG16+VGG19 enables the detection of both low-level and high-level features, whereas dropout, augmentation, and early stopping contribute to increasing the robustness and generalization.

5. Conclusion And Future Scope

The proposed research suggests the implementation of a powerful hybrid deep learning model that will be used to distinguish Intracranial Hemorrhage (ICH) based on CT scan images, in a way that will be efficient in capturing both low-level and high-level features to present a multi-scale and highly discriminative description of hemorrhagic regions. The large preprocessing, data boosting, dropout and early stopping provide model generalization and overfitting avoidance whereas the right dataset curation guarantees balanced representation of classes. The framework was evaluated on a curated dataset of 3,000 scans and reported 97.2% accuracy, 96.8% sensitivity and 97.6% specificity, 97.0% F1-score and 0.985 AUC, surpassing the state-of-the-art CNN-based, hybrid, and conventional machine learning methods, and thus demonstrates a wide clinical potential as a decision support tool. In perspective, the framework can be improved with multi-modal imaging, explainable AI to achieve interpretability, 3D volumetric analysis, real-time implementation in the emergency environment, inter-institutional validation for broader applicability, and multi-task learning for joint detection and localization of hemorrhages. Altogether, this hybrid VGG16/VGG19 model forms a good base of AI-based neuroimaging and the further development of this methodology is likely to achieve higher accuracy, interpretability, and clinical applicability in the critical care setting.

References

1. Mohamed Khalifa and Mona Albadawy, "AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency", Computer Methods and Programs in Biomedicine Update, Vol. 5, pp. 1–12, 2024.
2. Yan Xu, Rixiang Quan, Weiting Xu, Yi Huang, Xiaolong Chen and Fengyuan Liu, "Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches", Bioengineering, Vol. 11, No. 10, pp. 1–42, 2024.
3. Anshu Mahajan and Ashima Mahajan, "Neuroimaging: CT Scan and MRI", Principles and Practice of Neurocritical Care, pp. 189–215, 2024.
4. A.A. Talal Abdullah, Mohd Soperi Mohd Zahid and Waleed Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges and Future Directions", Symmetry, Vol. 13, No. 12, pp. 1–9, 2021.
5. Marcel Heinrich, M.R.H. Ahmed Mostafa, P. Jennifer Morton, J.A.C. Lukas Hawinkels and Jai Prakash, "Translating Complexity and Heterogeneity of Pancreatic Tumor: 3D in Vitro to in Vivo Models", Advanced Drug Delivery Reviews, Vol. 174, pp. 265–293, 2021.
6. Quoc Tuan Hoang, Xuan Hien Pham, Xuan Thang Trinh, Anh Vu Le, V. Minh Bui and Trung Thanh Bui, "An Efficient CNN-based Method for Intracranial Haemorrhage Segmentation from Computerized Tomography Imaging", Journal of Imaging, Vol. 10, No. 4, pp. 1–15, 2024.
7. Payal Malik, Ajay Dureja, Aman Dureja, Rajkumar Singh Rathore and Nisha Malhotra, "Enhancing Intracranial Haemorrhage Diagnosis through Deep Learning Models", Procedia Computer Science, Vol. 235, pp. 1664–1673, 2024.
8. Cansu Yalcin, Valeriia Abramova, Mikel Terceno, Arnau Oliver, Yolanda Silva and Xavier Llado, "Hematoma Expansion Prediction in Intracerebral Haemorrhage Patients by using Synthesized CT Images in an End-to-End Deep Learning Framework", Computerized Medical Imaging and Graphics, Vol. 117, pp. 1–9, 2024.
9. Anandakumar Haldorai, Suriya Murugan and Minu Balakrishnan, "Haemorrhage Detection from Whole-Body CT Images using Deep Learning", Artificial Intelligence for Sustainable Development, pp. 139–151, 2024.
10. Andrea Zirn, Eva Scheurer and Claudia Lenz, "Automated Detection of Fatal Cerebral Haemorrhage in Postmortem CT Data", International Journal of Legal Medicine, Vol. 138, No. 4, pp. 1391–1399, 2024.
11. Chi-Tung Cheng, Chun-Hsiang Ooyang, Chien-Hung Liao and Shih-Ching Kang, "Applications of Deep Learning in Trauma Radiology: A Narrative Review", Biomedical Journal, Vol. 48, No. 1, pp. 1–7, 2025.
12. A.S. Neethi, Santhosh Kumar Kannath, Adarsh Anil Kumar, Jimson Mathew and Jeny Rajan, "A Comprehensive Review and Experimental Comparison of Deep Learning Methods for Automated Haemorrhage Detection", Engineering Applications of Artificial Intelligence, Vol. 133, pp. 1–9, 2024.
13. Suleyman Uzun and Mehmet Okuyar, "A New Deep Learning-based GUI Design and Implementation for Automatic Detection of Brain Strokes with CT Images", The European Physical Journal Special Topics, Vol. 234, No. 1, pp. 141–164, 2025.
14. Abdulkader Helwan, Georges El-Fakhri, Hadi Sasani and Dilber Uzun Ozsahin, "Deep Networks in Identifying CT Brain Hemorrhage", Journal of Intelligent & Fuzzy Systems, Vol. 35, No. 2, pp. 2215–2228, 2018.
15. Hamza Sekkat, Abdellah Khallouqi, Omar El Rhazouani and Abdellah Halimi, "Automated Detection of Hydrocephalus in Pediatric Head Computed Tomography using VGG16 CNN Deep Learning Architecture and Automated Segmentation Workflow for Ventricular Volume Estimation", Journal of Imaging Informatics in Medicine, pp. 1–18, 2025.
16. Omid Mirzaei, Sedra Alialsagher Mohammed, Boran Sekeroglu and Ahmet Ilhan, "Comparison of Intracranial Haemorrhages Detection Performances of Deep Learning Models on CT Images", Procedia Computer Science, Vol. 258, pp. 3194–3202, 2025.
17. Simarjeet Kaur and Amar Singh, "A New Deep Learning Framework for Accurate Intracranial Brain Haemorrhage Detection and Classification using Real-Time Collected NCCT Images", Applied Magnetic Resonance, Vol. 55, No. 6, pp. 629–661, 2024.
18. Yurui Hu, Tianyu Liu, Shutong Pang, Xiao Ling, Zhanqiu Wang and Wenfei Li, "Deep Learning-Assisted Diagnosis of Placenta Accreta Spectrum using the DenseNet-121 Model: A Multicenter, Retrospective Study", Journal of Imaging Informatics in Medicine, pp. 1–10, 2025.
19. Mohemmed Sha, "A Graph Neural Network Technique for the Prediction of Cerebral Stroke using an Unbalanced Medical Dataset", Multimedia Tools and Applications, pp. 1–37, 2025.
20. Yu-Ruei Chen, Chih-Chieh Chen, Chang-Fu Kuo and Ching-Heng Lin, "An Efficient Deep Neural Network for Automatic Classification of Acute Intracranial Haemorrhages in Brain CT Scans", Computers in Biology and Medicine, Vol. 176, pp. 1–11, 2024.
21. Ding, W.; Zhou, T.; Huang, J.; Jiang, S.; Hou, T.; Lin, C.T. FMDNN: A Fuzzy-Guided Multi-Granular Deep Neural Network for Histopathological Image Classification. IEEE Transactions on Fuzzy Systems, 2024.

22. Zhang, L.; Wang, Q.; Zhao, S.; Liu, D.; Li, C.; Wang, B.; Wang, X. Deep-Learning-Based Microwave-Induced Thermoacoustic Tomography Applying Realistic Properties of Ultrasound Transducer. IEEE Transactions on Microwave Theory and Techniques, 2024.
23. Chagahi, M.H.; Piran, M.J.; Delfan, N.; Moshiri, B.; Parikhan, J.H. AI-Powered Intracranial Hemorrhage Detection: A Co-Scale Convolutional Attention Model with Uncertainty-Based Fuzzy Integral Operator and Feature Screening. arXiv preprint arXiv:2412.14869, 2024.
24. He, B.; Xu, Z.; Zhou, D.; Zhang, L. Deep Multiscale Convolutional Feature Learning for Intracranial Hemorrhage Classification and Weakly Supervised Localization. Heliyon, 2024, 10.
25. Ragab, M.; Salama, R.; Alotaibi, F.S.; Abdushkour, H.A.; Alzahrani, I.R. Political Optimizer with Deep Learning-Based Diagnosis for Intracranial Hemorrhage Detection. IEEE Access, 2023, 11, 71484–71493.
26. Negm, N.; Aldehim, G.; Nafie, F.M.; Marzouk, R.; Assiri, M.; Alsaid, M.I.; Drar, S.; Abdelbagi, S. Intracranial Haemorrhage Diagnosis Using Willow Catkin Optimization with Voting Ensemble Deep Learning on CT Brain Imaging. IEEE Access, 2023.
27. Malik, P.; Vidyarthi, A. A Computational Deep Fuzzy Network-Based Neuroimaging Analysis for Brain Hemorrhage Classification. IEEE Journal of Biomedical and Health Informatics, 2023.