MEKATRONIKA
JOURNAL OF MECHATRONICS AND INTELLIGENT MANUFACTURING
E-ISSN: 2637-0883
VOLUME 6, ISSUE 2, JULY, 2024, 20 – 25
DOI: https://doi.org/10.15282/mekatronika.v6i2.11358

RESEARCH ARTICLE

Keras Implementation in Detecting Intracranial Hemorrhage and Multiclass Classification of Subtypes via Transfer Learning and Classifiers Selection

Amiir Haamzah Mohamed Ismail[1] and Anwar P.P. Abdul Majeed[2,*]

[1] Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
[2]Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia

ABSTRACT - The development of deep neural networks for medical imaging applications, especially the diagnosis of intracranial hemorrhage (ICH) from CT scans, is greatly aided by machine learning frameworks such as Keras. This work investigates a pipeline that uses Keras' neural modules to distinguish between CT scans of the normal head and those with ICH. Transfer learning models are then used to categorize ICH subtypes. An extensive analysis of current research and techniques demonstrates the effectiveness of deep learning in medical imaging and emphasizes how AI may improve radiologists' diagnostic precision. Using windowing techniques to improve diagnostic features, the study preprocesses pictures from the RSNA Intracranial Hemorrhage Detection dataset. The study assesses performance indicators such classification accuracy using SVM, k-NN, and Random Forest classifiers combined with built-in models from Keras, such as Xception and DenseNet. Findings show that the Xception-SVM pipeline performs exceptionally well in binary classification tasks, achieving 76.33% accuracy, while DenseNet201-SVM performs well in multiclass classification, achieving 60% accuracy. These results highlight how crucial it is to choose the right pipelines for certain classification jobs in order to achieve the best results possible when using medical image analysis. In order to improve diagnostic precision in identifying cerebral hemorrhages, future research directions include increasing classifier performance, investigating sophisticated preprocessing techniques, and fine-tuning models.

ARTICLE HISTORY
Received : 25th June 2024
Revised : 4th July 2024
Accepted : 6th Sept 2024
Published : 30th Sept 2024

KEYWORDS
Intracranial Hemorrhage
CT scan
Keras
Classifier

1.0 INTRODUCTION

Machine learning academics are familiar with Keras, an AI framework developed by Francois Chollet for rapid testing and prototyping in deep neural networks. Keras enables deep learning models for illness diagnosis, prognosis, and therapy prediction based on biomedical data such as genomics, transcriptomics, proteomics, and imaging in which that help physicians make correct and fast judgments, which leads to better patient outcomes (1).

CT scans are the primary imaging modality used to diagnose intracranial hemorrhage (ICH) (2). This form of diagnosis is chosen because of the nature of CT scans, which identify bone, blood, and brain tissues based on their densities (3). A visual assessment, performed by doctors, radiologists, or in-house residents, will determine the afflicted region, level of damage, and perhaps the origin and source of the bleeding.

As a result, several articles claim that artificial intelligence (AI) can be used in medical imaging to help with illness identification. AI can help radiologists detect ICH more accurately and identify ICH subtypes, reducing misdiagnosis and improving patient outcomes (4).

The general goal of this study is to leverage the Keras neural modules to develop an AI pipeline that is able to identify normal CT scan images of the head compared to images that intracranial is hemorrhage is present. Subsequently, the type of intracranial hemorrhage is studied on what best AI best pipeline is able to properly detect them. About 17 built in Keras transfer learning models are evaluated in which they are passed on three state of the art classifiers for detection of the hemorrhage condition, namely support vector machine, k-nearest neighbors, and random forest classifier. This will also highlight which modality is suitable for implementing intracranial detection in CT scan images, through binary class classification of multiclass classification.

1.1 Related Work

Previous studies demonstrated the current advances and future promises of neuroimaging and deep learning for brain stroke detection. Encompassing 113 research publications published in databases such as IEEEXplore, PubMed, ScienceDirect, and Google Scholar. The studies were then sorted and filtered based on picture modality, datasets, technique, loss function, epoch count, optimizer, framework or library utilized, and validation system. The findings imply that a transfer learning strategy may be used to train the model, and data augmentation can help prevent overfitting. This

*CORRESPONDING AUTHOR | Anwar P. P. Abdul Majeed | ✉ anwarm@sunway.edu.my

is to emphasize that further study is needed to guarantee that deep learning can be used properly for medical picture diagnosis (5).

By reviewing current deep learning techniques for the automatic identification of acute cerebral hemorrhage and its subtypes, a simulation for cerebral bleeding detection using the RSNA dataset. The acquired images were preprocessed using techniques such image augmentation and windowing, whereby the window width (WW) and window level (WL) were adjusted. The characteristics from the CT scan pictures are then extracted using VGG16, and CNN is used for classification (6).

Timely reporting of crucial results in radiology can improve patient outcomes. The goal of the study was to examine the effects of Aidoc, a well-known machine learning algorithm that has been authorized by the US Food and Drug Administration (US FDA), on radiologists' workflow quality. In order to conduct the study, the scientists extracted 30,124 noncontrast head CT scans using data from the Radiology Information System. As a result, the Aidoc framework's feature extraction, learning model, and classifier are confidential and were not shared in this study. Nonetheless, the experiment's outcomes demonstrate that the length of stay (LOS) and report turnaround time (RTAT) of ICH patients can be shortened in these two areas (7).

2.0 METHODS AND MATERIAL

Following a review of the literature and other studies, it was established that this study is carried out in five basic stages. Phase 1 involves gathering data, Phase 2 involves preprocessing it, Phase 3 involves extracting features using the transfer learning (TL) technique, Phase 4 involves classifying images, and Phase 5 involves evaluating data-centric robustness and choosing the optimal model. After ICH is detected and classified into the subtypes of intraparenchymal (IPH), intraventricular (IVH), subarachnoid (SAH), subdural (SDH), and epidural (EDH), the optimal pipeline can be suggested based on its performance metrics.

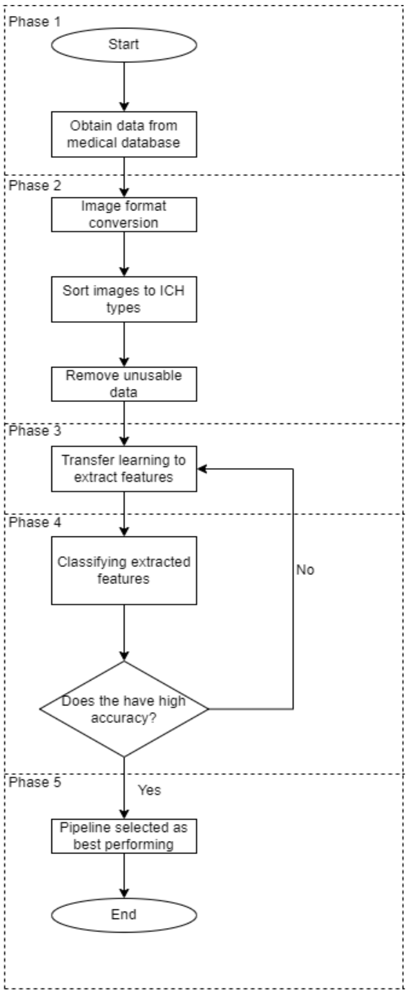

Figure 1. Research workflow

The RSNA Intracranial Hemorrhage Detection dataset was chosen for this investigation. This dataset's imaging modality is CT with a 5 mm slice thickness that isn't contrast enhanced. In addition, the image file has a tags-based structure specified in the DICOM standard. The labeler demographics were then 60 neuroradiologists who were members of the ASNR, a mix of junior and senior members. Inconsistent labels were eliminated from the data or rectified throughout the adjudication procedure. As a result, the three institutions that annotated the images guaranteed the accuracy and confidentiality of the history of patient illness (HPI) and identified instances in which the imaging plane, body part, or quality was incorrect. When it comes to HPI, all information about age and gender is hidden, thus the dataset only includes visual cues from CT scan images.

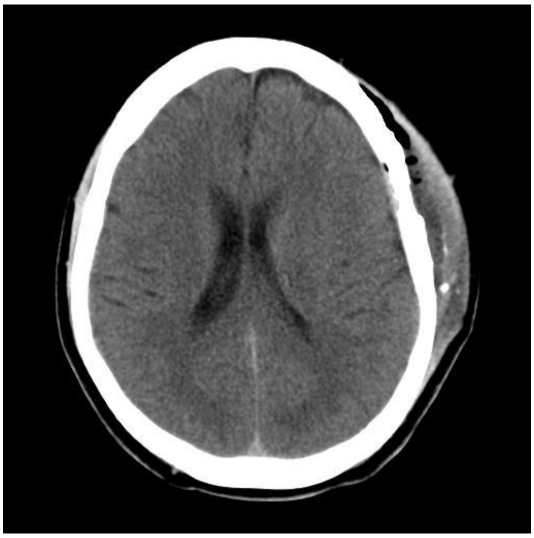

Figure 2. Sample from dataset

The format in which CT scans are saved varies based on the manufacturer's scanner settings. Although there are other formats, DICOM and NIfTI are the most often used ones. The Digital Imaging and Communications in Medicine (DICOM) format used for the RSNA dataset comprises a header and image data that are placed in a single file.

By configuring the brain windows to W:40 L:80, windowing is applied to the pictures in order to highlight the most common aspects of the CT scans. These are the Hounsfield units (HU), expressed as level and width (W:x L:y). Window level/center (WL) is the midpoint of CT numbers and acts as an image brightness increaser when CT number lowered and vice versa. Window width (WW) is the range of CT numbers functioning as contrast adjuster (8). The original image, represented by the Hu image in Figure 3.3, is the category of setting, which corresponds to the usual values used in CT scan imaging.

Subsequently, corrupted or unreadable photos are removed through data cleaning in order to eliminate low quality data or data with inadequate features. Subsequently, the data is sorted using a Python script into balance groups of healthy and suspected ICH. Extrapolating from the dataset into intraparenchymal (IPH), intraventricular (IVH), subarachnoid (SAH), subdural (SDH), and epidural (EDH) spaces would result in another balance dataset. Table 2.1 lists each category. The data is split 70:30 for the train and validation sets, and the validation sets are further separated into 50:50 ratios for testing and validation, respectively.

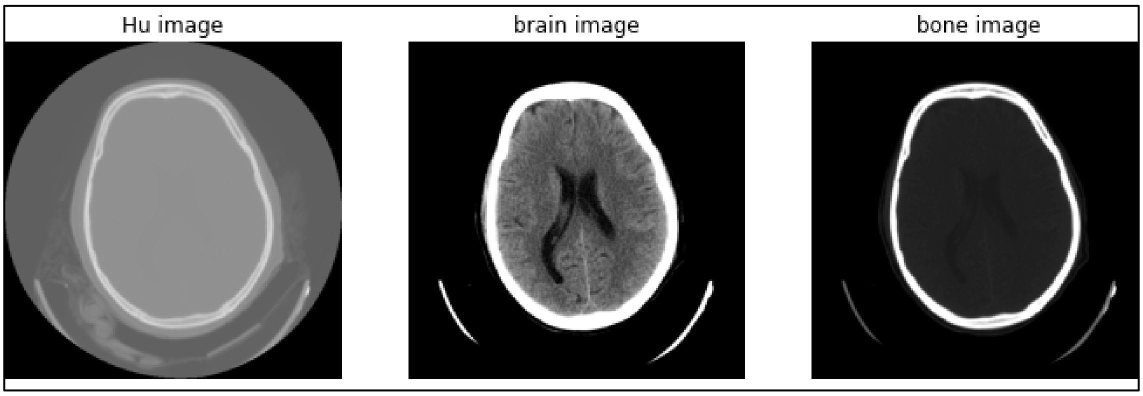

Figure 3. Windowing effects on CT scan images

There is no mention of the machine learning libraries or frameworks used in the experiments in the literature examined in the preceding studies. It is crucial to understand which libraries are being utilized because they offer varying machine learning tools and models. In this experiment, the default machine learning models from the Keras machine learning library are used. These models employed are shown in table 1.

Table 1. Built in Keras models

Model	Size (MB)	Parameters
Xception	88	22.9M
VGG16	528	138.4M
VGG19	549	143.7M
ResNet50	98	25.6M
ResNet50V2	98	25.6M
ResNet101	171	44.7M
ResNet101V2	171	44.7M
ResNet152	232	60.4M
ResNet152V2	232	60.4M
InceptionV3	92	23.9M
InceptionResNetV2	215	55.9M
MobileNet	16	4.3M
MobileNetV2	14	3.5M
DenseNet121	33	8.1M
DenseNet169	57	14.3M
DenseNet201	80	20.2M
NASNetMobile	23	5.3M

The features that were retrieved from the CT scans will be integrated into the most advanced machine learning algorithms. This happens after the tested transfer learning model extracted the features and was assessed by Random Forest (RF), k-Nearest Neighbor (k-NN), and Support Vector Machine (SVM) classifiers. As part of supervised machine learning (ML), the Support Vector Machine (SVM) technique performs hyper-plan segregation of the data with the highest margin and then uses the kernel trick to transform the input datasets into a higher dimensional space (9). This classifier is often used for binary classification problems in machine learning due to its ability to handle big datasets. A well-known and conventional machine learning technique is k-Nearest Neighbor (k-NN). Known as the traditional classification, the nonparametric conventional supervised classification approach (also called the k-NN) yields a desirable classification accuracy (CA) for the superior k value parameter. A similarity component known as a distance function is necessary for k-NN to maintain all of the previously categorized examples and classify new cases. Pattern recognition and mathematical estimation are two common applications for k-NN. Random Forest (RF) is an ensemble machine learning classification strategy developed by merging several decision trees during model training and categorization (10). It should be emphasized that RF is composed of numerous distinct decision trees that collaborate to construct a categorizing model. An RF's structure is comprised of two groups of trees. It combines a decision tree with a bagging classifier. While increasing the trees, RF introduces additional uncertainty. The program searches for the standout attribute among a random set of qualities. This search option produces a wide range of results, often resulting in a higher model classification accuracy (CA).

The randomized search technique algorithm was utilized to fine-tune the hyperparameters of each classifier. It is important to note that this study has two distinct classification goals: the multiclass categorization of ICH subtypes and the binary classification of healthy and suspected ICH cases. The process of choosing the most effective collection of parameters for machine learning algorithms is known as hyperparameter tuning. Nevertheless, putting the one-value-at-a-time (OVAT) technique into practice is a very laborious operation that might not always produce the greatest outcomes. The grid search method and the random search method are hence the options. The grid search method analyzes the grid's properties and chooses the combination that yields the highest yield. However, when the algorithm thoroughly explores every conceivable combination, this comes at the expense of time and computational resources. In contrast, the randomized search approach samples each parameter setting from a distribution of possible values by conducting a randomised search over the parameters. The benefits of this approach are that parameters can be chosen irrespective of the quantity of parameters or their potential values, and efficiency is unaffected by the inclusion of features that don't impact performance. It is important to note that since this research does not cover the fine tuning of every parameter value accessible, the values for the parameters for each classifier are chosen based on their default state in the Keras library.

The model's performance in carrying out classification tasks can be determined using a variety of performance indicators. Because the balance dataset uses methods that can produce unbiased predictions, the classification accuracy (CA) is employed in this work to test the machine learning (ML) pipelines in recognizing the existence of ICH in the CT scan images. Equation (1) provides the formula for classification accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where, TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.

3.0 RESULTS AND DISCUSSION

Analysis of binary classification using transfer learning (TL) models and the KNN classifier showed that different architectures performed differently. Although certain models demonstrated elevated training accuracy, their capacity to extrapolate novel data differed. As the best performer, MobileNetV2 demonstrated how well it could capture complex features. The findings highlight the necessity for models that perform well during training and efficiently generalize to new situations. Additional insights were obtained by exploring TL models in binary classification using the SVM classifier. Test accuracy was higher for models with great generalization, such as DenseNet201 and MobileNet, than for other models. Not to be outdone, Xception showcased the finest of both model design and classifier selection. The results highlight how crucial it is to assess TL models using several classifiers in order to achieve the best results possible when classifying medical images. When the Random Forest classifier was used for binary classification, different model performances were shown in the findings. Strong generalization was shown by MobileNet, DenseNet169, and DenseNet201, although VGG16 and ResNet50 had difficulties. The classifier selection was important, and it was clear that the Random Forest worked well for using TL characteristics. For practitioners looking to optimize TL models for binary classification tasks, the research offered insightful information.

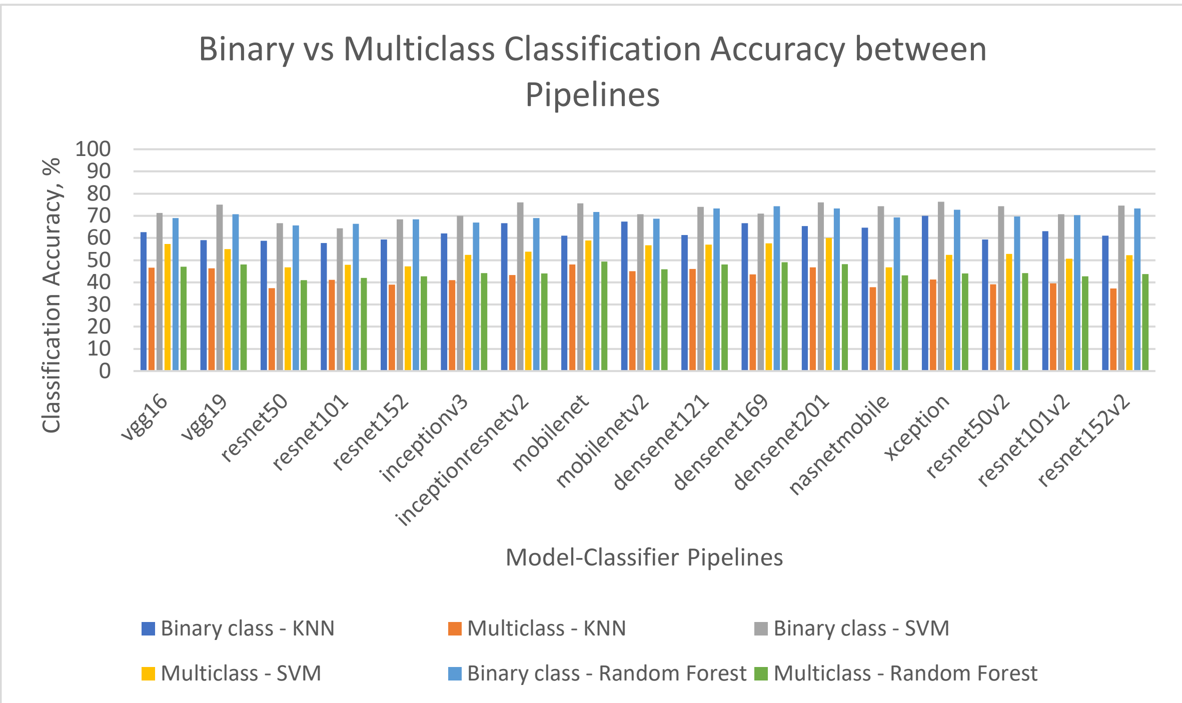

Figure 4. Experiment Results

The careful procedure of dataset optimization and classifier selection was essential for the multiclass categorization of ICH types. A diversified dataset was required, and computational restrictions were balanced by including 1000 images in each class. Strong generalization was demonstrated by MobileNet and DenseNet201, which were found to be promising options. The experiment also brought attention to persistent issues including dataset heterogeneity and class imbalance, highlighting the necessity of constant improvement and optimization. MobileNet distinguished itself in the field of multiclass classification using the KNN classifier by achieving the greatest testing accuracy, demonstrating its capacity to identify a wide range of patterns. The outcomes highlighted how difficult it is to achieve high accuracy in this challenging endeavor. Xception, on the other hand, was the best performance as identified by the SVM classifier, demonstrating its ability to extract and generalize features efficiently. DenseNet201 demonstrated strong performance as well, highlighting the necessity of closely balancing the demands of classifier performance, processing capacity, and dataset features.

The study concludes that, with a classification accuracy of 76.33%, the XceptionSVM pipeline is the most efficient method for binary classification jobs. However, the DenseNet201-SVM pipeline achieves a 60% classification accuracy, outperforming competing models for multiclass classification problems. These results emphasize the significance of choosing the right pipeline based on the particular classification task in order to optimize performance and accuracy.

The study's findings provide potential routes for future investigation. Model refinement could benefit from further investigation, including sophisticated pre-processing approaches and innovative transfer learning architectures. Furthermore, understanding the impact of hyperparameter tuning and optimizing classifier performance may lead to more accurate and reliable ICH detection models. Essentially, the goals of the study are emphasized in the conclusion, which emphasizes the need for ongoing development and improvement in the field of medical picture analysis.

Advancements in technology and complex datasets require continual study and development to provide accurate diagnostic tools for healthcare practitioners detecting intracranial hemorrhages.

4.0 CONFLICT OF INTEREST

The authors declare no conflicts of interest.

5.0 REFERENCES

[1] Barillaro L. Deep Learning Platforms: Keras. Reference Module in Life Sciences. 2024 Jan 1;

[2] Intracranial hemorrhage | Radiology Reference Article | Radiopaedia.org [Internet]. 2022 [cited 2023 Oct 9]. Available from: https://radiopaedia.org/articles/intracranial-haemorrhage

[3] Cerebrovascular Disease – Classifications, Symptoms, Diagnosis and Treatments [Internet]. [cited 2023 Oct 9]. Available from: https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Cerebrovascular-Disease

[4] Warman R, Warman A, Warman P, Degnan A, Blickman J, Chowdhary V, et al. Deep Learning System Boosts Radiologist Detection of Intracranial Hemorrhage. Cureus [Internet]. 2022 Oct 13 [cited 2023 Oct 9];14(10). Available from: /pmc/articles/PMC9653089/

[5] Karthik R, Menaka R, Johnson A, Anand S. Neuroimaging and deep learning for brain stroke detection - A review of recent advancements and future prospects. Comput Methods Programs Biomed. 2020;197.

[6] Rane H, Warhade K. A survey on deep learning for intracranial hemorrhage detection. 2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021. 2021;38–42.

[7] Davis MA, Rao B, Cedeno PA, Saha A, Zohrabian VM. Machine Learning and Improved Quality Metrics in Acute Intracranial Hemorrhage by Noncontrast Computed Tomography. Curr Probl Diagn Radiol. 2020;000:6–11.

[8] Baba Y, Murphy A. Windowing (CT). Radiopaedia.org [Internet]. 2017 Mar 22 [cited 2024 Jul 1]; Available from: http://radiopaedia.org/articles/52108

[9] Gholami R, Fakhari N. Support Vector Machine: Principles, Parameters, and Applications. 1st ed. Handbook of Neural Computation. Elsevier Inc.; 2017. 515–535 p.

[10] Louppe G. Understanding Random Forests: From Theory to Practice. 2014 Jul 28 [cited 2023 Oct 9]; Available from: http://arxiv.org/abs/1407.7502