An Attention-Based ResNet Architecture for Acute Hemorrhage Detection and Classification: Toward a Health 4.0 Digital Twin Study

AFTAB HUSSAIN1, MUHAMMAD USMAN YASEEN1, MUHAMMAD IMRAN1, MUHAMMAD WAQAR1, ADNAN AKHUNZADA2, (Senior Member, IEEE), MOHAMMAD AL-JA’AFREH3,4, AND ABDULMOTALEB EL SADDIK3,4

1Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 45550, Pakistan
2College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar
3School of Electrical Engineering and Computer Sciences, University of Ottawa, Ottawa, ON K1N6N5, Canada
4Mohamed bin Zayed University of Artificial Intelligence, Masdar City, Abu Dhabi, United Arab Emirates

Corresponding author: Mohammad Al-Ja’afreh (jaafreh@uottawa.ca)

ABSTRACT Due to the advancement of digital twin (DT) technology, Health 4.0 applications have become reality and starting to take roots. In this article, we focus on intracranial hemorrhage (ICH) which is a life-threatening emergency that needs immediate diagnosis and treatment. ICH is caused by bleeding inside the skull or brain. Radiologists typically examine computed tomography (CT) scans of the patients to determine the ICH and its subtype. But the manual assessment of the CT scan is a complex and time-consuming task. The existing pre-trained convolutional neural network (CNN) models are state-of-the-art for ICH classification. However, they employ poor feature extraction techniques which hinder overall model performance. Furthermore, they suffer from the curse of dimensionality and use redundant and noisy features. The problem of imbalanced data is also crucial for achieving model generalization. This paper proposes a hybrid attention-based ResNet architecture for ICH detection and classification. An attention mechanism allows the model to focus on a specific region and extract relevant features. Principal component analysis (PCA) is used for dimensionality reduction and redundant feature removal whereas deep convolutional generative adversarial network (DCGAN) is used for resolving the class imbalance problem. The proposed model is evaluated using the dataset assembled during the Radiologist Society of North America (RSNA) ICH detection challenge 2019. The results show that our proposed model outperforms existing state-of-the-art models in terms of accuracy and F1-score. ICH classification achieved accuracies of 99.2%, 97.1%, 96.7%, 96.7% and 96.1%, for detecting epidural hemorrhage (EH), intraparenchymal hemorrhage (IH), intraventricular hemorrhage (IVH), subdural hemorrhage (SH), and subarachnoid hemorrhage (SAH) subtypes respectively. The F1-score of 96.1% for EH subtype is also best when compared with the benchmark models.

INDEX TERMS Computed tomography, data augmentation, deep convolutional generative adversarial network, deep learning, digital twin, health 4.0, intracranial hemorrhage, principal component analysis, ResNet-152V2.

I. INTRODUCTION

AI in healthcare is getting significant attention due to new challenges faced by caregivers, patients, and providers [1]. The mastership to simulate, experience and adjust the real-world environment based on learning from working on its digital replica played a huge role in helping scientists and engineers improve their success rate. It enhances their understanding of an equipment/product/process and predictability of positive outcome in various domains [1]. With the merits of machine learning engineering, the success achieved by stakeholders in other verticals, serves as a motivation
for patient advocates and researchers, to reduce risk and 
cost of healthcare. The potential of implementing Health 4.0 
technology, via the incorporation of artificial intelligence 
solutions, to improve the quality of service in healthcare is 
tremendous [2], [3].

Developing a digital solution for intracranial hemorrhage 
(ICH) detection is part of the Health 4.0 initiative. ICH refers 
to the outflow of blood from a cranial blood vessel. It may 
be caused by both intrinsic and extrinsic factors. Confusion, 
sudden headache, dizziness, nausea, vision loss or barely see-
ing, dyspnea, and vomiting are symptoms of this disease [4]. 
The human brain is composed of three meninges known as 
the pia mater (PM), dura mater (DM), and arachnoid (AD), 
which are located between brain tissues and the skull. 
The distinct types of brain hemorrhages can be identified 
based on bleeding location in the brain tissues. In cases 
of epidural hemorrhage (EH), blood seeps from the skull 
into the DM. Subarachnoid hemorrhage (SAH) is caused by 
bleeding inside the subarachnoid. Subdural hemorrhage 
(SH) is the term used to describe the bleeding caused by 
AD meningitis and DM. Intraparenchymal hemorrhage (IH) 
refers to the bleeding within the brain’s parenchyma, whereas 
intraventricular hemorrhage (IVH) pertains to blood outflow 
in the brain’s ventricular system [5]. The frequency of ICH 
is estimated to be 10–30 per 100,000 people worldwide [5]. 
Whereas in US, 67,000 ICH cases are reported per year, with 
a 30-day mortality rate ranging from 35% to 52% [6]. ICH 
can severely damage the brain quickly, endangering human 
life, and is implicated in two million strokes worldwide [6]. 
Although, ICH strokes account for only 10-15% of all strokes 
but have a mortality rate of 50% [5].

ICH is a critical disease that can result in permanent mental 
disability or death. It necessitates an emergency diagnosis and 
treatment. CT scans are a non-invasive and reliable imaging 
system for diagnosing ICH. Fig. 1 shows the sample non-
contrast CT scan images which are widely used in ICH 
detection and subtype classification. When patients exhibit 
nervological ICH symptoms, CT images are taken and exam-
ined by a competent radiologist for ICH detection. A radiolo-
gist’s manual examination of a CT scan is a time-consuming 
and challenging task [7]. The need for emergency treatment, 
combined with critical conditions, may result in the patient’s 
death. As a consequence, an automatic system to detect ICH 
and its subtype is required to contend with critical conditions, 
diagnostic time, and assistance of radiologist. The perform-
ance of the reasoned automated tool in medical diagnosis 
should be high. The brain tissue from the various types of ICH 
differs only slightly, making it difficult to distinguish among 
different ICH subtypes.

Researchers have proposed a range of machine learning 
and deep learning models for ICH detection and classification. 
These models either perform binary classification, which 
only differentiates between healthy and hemorrhagic images, 
or multiclass classification to distinguish between different 
ICH subtypes [8], [9]. Feature selection is a complicated 
task that requires more attention to accurately select the 
required features. The existing pre-trained convolutional neu-
networks (CNNs) models are cutting-edge for multiclass 
image classification. The motivation for the field of medical 
imaging is pre-trained CNNs trained on large datasets such as 
ImageNet etc.

The existing CNN models for ICH classification have 
many limitations. They use poor feature extraction techniques 
which select insufficient features [7], [10], [11], [15], [20]. 
Furthermore, the curse of dimensionality during feature 
extraction results in high computational overhead [7], [10], 
[11], [15], [20], [22]. The inclusion of poorly balanced and 
denormalized data is another limitation for model general-
ization in multiclass classification [4], [5], [10], [15]. Small 
datasets are also the primary cause of model underfitting in 
supervised learning [11], [15], [20], [22].

In this paper, we propose a novel deep learning-based CNN 
model to efficiently detect and classify brain hemorrhage 
and its subtypes. The proposed model has an attention-based 
mechanism for feature extraction, DCGAN to augment the 
data, PCA for feature selection, and XGB classifier to classify 
the ICH subtypes on CT scans. The primary contributions 
of this work are listed below.

1) We present an attention-based ResNet-152V2 
arquitecture to address the issue of insufficient spatial 
and intensity-related feature extraction. Our pro-
posed hybrid feature extractor extracts pertinent, 
latent, complex, and hidden features from the high-
dimensional feature map, improving model computa-
tional efficiency.

2) A deep convolutional generative adversarial network 
(DCGAN)-based data augmentation model is used to 
solve the class imbalance problem in the RNSA-2019 
dataset. We generate CT images for EH minority class 
with close resemblance to the real CT scan images.

3) We applied principal component analysis (PCA) to 
overcome the challenge of feature selection, redundant 
feature removal, and dimensionality reduction. We only 
retain the 90 features with an explained variance of 
PCA from the 2048 input feature vector and discard the 
others.

The remaining paper is organized as follows. Section II 
explores the literature related to ICH detection and classi-
fication. Section III elaborates on different components of 
the proposed model. The experimental results are discussed 
in Section IV. Section V presents a comparative analysis of 
the proposed work with existing benchmark techniques. 
Finally, Section VI provides conclusion of the paper and 
future research work directions.

II. RELATED WORK

There are many research studies in the fields of medicine, 
machine learning and deep learning, that target CT scans 
assessment to detect and classify acute brain hemorrhage, 
a critically important task in patient care. These studies can
be classified based on the techniques used for implementing the model. Fig. 2 presents a taxonomy for classification of these techniques.

First group of studies are the traditional hardware/software/hybrid solutions from the medical field. They along with machine learning models such as support vector machine (SVM), random forest (RF) classifier etc make up the conventional approaches for ICH detection and classification. For example, in [16], authors used a microwave technology based device in conjunction with diagnostic mathematical algorithm for traumatic ICH detection in suspected brain hemorrhage patients. The model was trained on only 20 patients’ data, hence it is incapable of accurately predicting the novel variants of ICH. In [17], authors conducted a random forest (RF)-based study for ICH detection on CT scans. The key finding of this study is the dice similarity index (DSI) of 0.899. However, the CT images used in this model are manually segmented, which makes ICH detection in critical conditions unreliable and time-consuming. The work by [18] is a case study about ICH detection using a portable near-infrared device on Chinese population. The device only detects brain bleeding and can not differentiate between ICH subtypes. The model was trained on a dataset comprising CT scan and magnetic resonance imaging (MRI) images. The sensitivity and specificity of the algorithm were 95.6% and 92.5% respectively, which is also quite low when analysed for system accuracy.

Recently, the application of deep learning for ICH detection has increased due to superior performance of these models. For example, authors in [5] proposed ensemble learning algorithm using 2D CNN model and two sequence models for ICH detection and subtype classification. The log loss was 0.054, the area under the curve (AUC) was 0.944, the specificity was 0.988, and the sensitivity was 0.950. The model outperforms other existing models. However, by leveraging the attention mechanism, the feature extraction technique can further improve the performance of the model. The study in [19] uses deep convolutional neural network for detecting ICH using CT scan dataset. The AUC score of the model is 0.846 while the overall specificity and sensitivity are 0.800 and 0.730 respectively. The model limitation is that it only performs binary classification and cannot distinguish between ICH subtypes. The [20] proposes a model for detecting intracranial hemorrhage with deep learning using CNN (U-Net) architecture, The database of this study contains 43,00 CT scan images, which are insufficient to train the model. The study conducted by [21] uses 2D and 3D hybrid deep convolutional neural networks for hemorrhage detection on head CT scan images. The model achieved accuracy, sensitivity and specificity metrics as 0.970, 0.951, and 0.973,
respectively on the test set. The major limitation of this study is that the proposed solution is not generalised, it performed testing on a limited dataset from single academic institution.

In [22], a novel deep learning model is proposed that uses pre-trained deep convolutional neural network on a small dataset. The model attained an AUC score of 0.933, sensitivity of 98% and specificity of 95% on the test data of 200 cases. However, an insufficient training dataset of only 904 patients is used by the model. The [23] performed ICH detection and segmentation on CT scans using a patch-based fully convolutional network (PFCN) model. The model achieved an AUC of 0.976 and 0.966 for retrospective and prospective test cases respectively. An insufficient dataset of 591 CT scans is used for training, resulting in poor generalisation of the results. The study by [24] classified ICH and its subtypes using automated detection with training and validation of deep learning model on CT scans. The results demonstrate that ICH had an AUC of 0.9194. The authors’ main contribution was the generation of the ICH dataset, as the proposed deep learning method had limitations.

The work by [25] performed hematoma detection on CT scan reports using a natural language processing (NLP) pipeline method and obtained the distinctive features of SH. The model had an 84–90% accuracy when cross-checked with the results of physicians. However, the proposed study only looks at subdural (SH) reports, whereas other types of data are essential for avoiding class dominance. In [26], authors described a cascaded CNN developed by combining GoogleNet and dual FCN-8s architecture for detecting, improving, and explaining sensitivity in intracranial brain hemorrhage. The model accuracy was 98.28%, sensitivity was 97.91%, segmentation accuracy was 80.19% while precision and recall were both 82.15%. However, the proposed scheme only performs binary classification. The work by [27] performs automatic classification of ICH radiological reports using deep learning algorithm. It employs 1D-CNN for feature extraction and long short-term memory (LSTM) for text vector representation. However, the majority of the dataset used in this study was unlabeled data, whereas only small data reports were labelled. The study by [28] proposed hybrid CNN and LSTM method for detecting ICH in CT scan images. The model obtained weighted log loss of 0.05289, which is a better performance. However, the feature selection mechanism in the proposed scheme used all features, including useless features, making resources burdened.

The [29] work is about ICH recognition in smart cities using an innovative feedforward neural network (FFCN) and SVM algorithms. The SVM accuracy was 80.67% and FFCN accuracy was 86.7%. Because of insufficient training, the proposed prediction model is less generalized. The [30] performed ICH analysis on CT scans, which uses machine learning software to prioritise ICH types. The system has a sensitivity, specificity and accuracy of 88.7%, 94.2% and 93.4% respectively. An insufficient data is used to train the model, resulting in a less generalized model. In [31], authors used an automated deep learning mechanism for ICH classification on head CT scan images. The authors proposed SE-ResNeXt50, SE-ResNeXt50 ensemble, and EfficientNet-B3. The class imbalance dataset is used to train the model, which results in the model’s class learning dominance in predicting ICH. The [32] proposed ICH localization using weakly-supervised deep learning method implemented with the CNN model ResNet-like architecture. The model employed RSNA-2019 and CQ500 datasets for the experimentation. The results obtained were a dice coefficient of 58.08% and an accuracy of 89.54%. The model’s overall performance is comparatively better but has a low accuracy which needs to be improved.

The study by [33] proposed automatic ICH segmentation on the head CT scans using hybrid of the Fuzzy c-mean
and a distance regularised level set evolution algorithm. However, the sensitivity of the approach is 68.43%, which is too small to detect the ICH. Furthermore, the model is tested on the CT scans of merely 20 patients dataset. In [34], authors’ work is about ICH optimum segmentation with an inception network-based deep learning model for intracranial brain detection. The achieved sensitivity was 93.56%, the specificity was 97.56%, and the accuracy was 95.06%. The dataset used to train and validate the model in this scheme is small. It only uses 82 patients’ data points, which is insufficient to train the data-hungry CNN model. Authors in [35] proposed a CNN-based model to identify false negative error reported by radiologists in ICH detection. The results revealed that 1.6% of the cases were misclassified by radiologists as non ICH whereas they were ICH positive. Furthermore, the dataset used for this study was small. A deep learning-ensembled network based on EfficientNet proposed by [36] is used to detect ICH and subtypes in non-contrast CT scans. The main findings are 95.7% accuracy and 85.9% sensitivity. However, the dataset used for the subtype classification is biased.

Based on the literature review we identified following research gaps. The majority of the literature uses either small dataset or a dataset which is biased against any certain ICH subtype. Some of the articles only consider binary classification, while the performance of those targeting multiclass classification needs improvement. Most of the articles use basic CNN or LSTM models whereas only few articles attempt using pre-trained models which are state-of-the-art in image classification. Furthermore, a very rare set of articles focus on reducing the training time, which is quite a big issue if the dataset used for the subtype classification is biased.

To address these research gaps, this study proposes a hybrid attention-based CNN model for ICH detection and classification. This study proposes a hybrid attention-based CNN model for ICH detection and classification. In the first step, the RNSA-2019 dataset is acquired and analysed. The second step performs data augmentation of the RNSA-2019 dataset and generates CT scan images for the minority EH class. The augmented dataset is then passed to the third step for data preprocessing operations. The forth step takes preprocessed data as an input and applies attention-based ResNet-152V2 architecture for feature extraction. The weighted scaled features generated by the attention layer are then sent to PCA for dimensionality reduction. In the last step, the XGBoost classifier is used to perform the classification. The following subsections elaborate on each component of the model.

### III. PROPOSED ATTENTION-BASED RESTNET ARCHITECTURE

This study proposes a hybrid attention-based CNN model for ICH detection and classification. Fig. 3 presents the detailed working of the proposed model. The first step is to acquire and analyse the RNSA-2019 dataset. The second step performs data augmentation of the RNSA-2019 dataset and generates CT scan images for the minority EH class. The augmented dataset is then passed to the third step for data preprocessing operations. The forth step takes preprocessed data as an input and applies attention-based ResNet-152V2 architecture for feature extraction. The weighted scaled features generated by the attention layer are then sent to PCA for dimensionality reduction. In the last step, the XGBoost classifier is used to perform the classification. The following subsections elaborate on each component of the model.

#### A. DATA PREPROCESSING

The RNSA-2019 dataset comprises 16-bit DICOM (digital imaging and communications in medicine) CT scan files with images of size 512 × 512 × 3. Each file contains CT scan pixel data and metadata, and can be displayed in 65536 grayscale levels. The CT images are converted into 224 × 224 × 3 image sizes and are saved as PNG image files [10]. The resizing of the images is done to make them compatible with the ResNet-152V2 architecture input size.

The data preprocessing operations remove irrelevant information from each CT scan sample. It starts by applying intensity windowing to the CT scan images. CT images are made up of window center/level (WC) and window width (WW). Hounsfield unit (HU) is used to measure the intensity of these windows. We considered three intensity windows: the subdural window (L = 100 and W = 200), bone window (L = 600 and W = 2800), and brain window (L = 40 and W = 80). The values above 2800 are discarded because they show pixels from bones and hard materials. Whereas, we need to extract pixels from brain soft tissue to define the region of interest (ROI). The images are then normalized between 0 and 1 using Otsu’s method [10]. The skull removal algorithm takes subdural-window image as an input and removes high intensity parts by replacing them with zero values. The final preprocessed image is made up of red (G), green (G), and blue (B) colour channels. The RGB images are then passed to the attention-based CNN model for feature extraction. The preprocessing steps are illustrated in Fig. 4.

The CT scan images need to be standardized before being fed to the neural network. When they are fed into the neural network without proper feature scaling, the gradient exploding issue arises. It results in increasing the computation overhead and reduces the neural network’s convergence rate. We use the max-min scaling technique to normalise the CT scan data in the range 0 to 1. Eq. 1 provides a mathematical formula to calculate the max-min normalization.

\[
S(t) = \frac{x(t) - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

where \( x(t) \) is the original value and \( S(t) \) is the normalised value. \( x_{\text{min}} \) and \( x_{\text{max}} \) are the minimum and the maximum values in the \( x(t) \) respectively.

#### B. DATA AUGMENTATION MODULE

Traditional data augmentation techniques such as rotation, flipping, cropping, and translation are limited and cannot achieve optimized model generalization. This study generates epidural hemorrhage (EH) CT images using DCGAN to address the class imbalance problem in the RNSA-2019 dataset. DCGAN is an extension of the traditional GAN [37] with an additional deep convolutional layers to generate 2D images. In DCGAN, the pooling layer is abandoned, instead, fractional-strided convolutions are introduced in the generator and strided convolutions are added in the discriminator. ReLU activation function is used for all layers of the generator except the output layer which uses tanh. Whereas in discriminator, LeakyReLU is used as an activation function for all layers. [38].

The deep convolutional GAN outperforms in supervised learning and is capable of more powerful feature
convergence. Hence, DCGAN-based augmentation technique is used for generating replica EH CT scan images in the supervised learning mode. We used a training dataset comprising 3145 EH CT image samples from the RNSA-2019 ICH dataset. The proposed data augmentation model is built from the ground up using the DCGAN architecture [38]. The generator (G) takes random noise as an input and generates fake images that resemble real CT scan images. The role of the discriminator (D) is to determine if the generated CT images are fake or real. The generated images along with the real CT samples are fed as an input to the D. The final output of D indicates whether the generated sample is fake or real, represented by 0 and 1 respectively. Eq. (2) presents the minmax objective function for training the DCGAN-based data augmentation model.

\[
\min_{G} \max_{D} V(G, D) = \mathbb{E}_{r \sim P_{\text{data}}(r)}[\log D(r)] + \mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z)))].
\]

where \(D(r)\) is the probabilistic estimate of the discriminator that the real CT image \(r\) is real while \(D(G(z))\) is the probabilistic estimate of the discriminator that the generated fake CT scan image \(G(z)\) is real. \(P_{\text{data}}(r)\) denotes the real distribution \(r\) of the CT scan samples and \(P_z(z)\) represents the latent noise distribution \(z\). \(\mathbb{E}_r\) and \(\mathbb{E}_z\) represent the expected values of the discriminator \(D\) and generator \(G\) respectively.

Fig. 5 describes the working of the proposed DCGAN-based data augmentation model. The input to the G is a random noise \(Z\) which is a vector of shape 32765 sent as an input to the fully connected dense layer. The vector is then reshaped to a 4D tensor using reshape layer. The Conv2DTranspose layer performs fractional-stride convolution for upsampling of the data. The conv2D+LeakyReLU combination achieves learning and convergence on the upsampled data. It can repeat \(N\) number of times unless output image attains the desired shape. The output is then flattened and sent to dense layer, which uses tanh activation function to generate a CT scan image of size 224*224*3. The generated CT scan image along with the real CT scan image is then fed to the discriminator \(D\). The dropout is applied for all iterations of the discriminator except the first one, hence it is missing from first convolution block. The Conv2D (performs strided convolution)+LeakyReLU+Dropout block repeatedly performs convolution operations. The fully connected dense...
FIGURE 4. Sample implementation to highlight steps of the proposed model.

layer receives reshaped 1D vector using flatten layer, which is then used by the sigmoid activation function for binary classification of CT scan image being real(1) or fake(0). The sample generated images are shown in Fig. 6.

C. FEATURE EXTRACTION

The attention-based ResNet (residual) architecture is used for feature extraction from the RGB CT scan images. ResNet is a CNN-based architecture to address the vanishing gradient problem in deep neural networks (DNNs). Deep learning models are data-hungry and need a large amount of data to learn efficiently. It is difficult to collect abundant medical imaging and real-world CT scan data to train the data-hungry model. CNN’s transfer learning approach is the de-facto standard to overcome this problem in computer vision (CV), especially image classification. The idea behind transfer learning is to train the model on a larger dataset like ImageNet and then transfer the learning and weights of the model to the small dataset. The main advantage of the CNN models over the conventional techniques is that they automatically focus on the region of interest (ROI) and extract the critical features without being biased. The input layers focus on simple features, colours, shades, bars, corners, edges, etc. In contrast, the more complex and deeper layers are used to detect the more complicated features, objects, complex patterns, and textures. In the domain of CV, CNN has already achieved state-of-the-art performance. There are many medical imaging tasks handled efficiently with pre-trained CNNs.

We used pre-trained weights of the ResNet-152V2 architecture, and multiplied custom attention layers with the loaded model. ResNet architecture is the representative deep convolutional neural network composed of identity and convolutional blocks with perfect performance in image classification and computer vision. The bigger depth of the neural network causes the vanishing gradient problem, and as a result, model performance decreases. However, ResNet architecture effectively handles the vanishing gradient problem [39]. In this work, ResNet-152V2 dense layer is removed, the custom attention layers are defined in the attention network and the output of RestNet is passed directly to the attention network. In the attention network, we added a 2D convolutional layer with stride size 1, padding the same, and ReLU is used as an activation function. The dropouts in the
The working of the attention-based ResNet-152V2 architecture is presented in Fig. 7. The output from the pre-processing module is an image of size 224 × 224 × 3, particularly transformed to make the data compatible with the ResNet input size. The ResNet architecture comprises a stack of convolutional blocks and identity blocks. The added attention layers and 15% of the ResNet model are trained on the CT scan image data. The proposed model is ran for 50 epochs configured with an early stopper. All the weights along the model are saved to further extract the features from the prepared dataset for classification. In addition, trained model weights are loaded as a feature extractor after the global average pooling layer feature vector is extracted. The flattened layer converts the feature map into a 1D vector, which will be sent as an input to the fully connected layer. Eq. (3) provides the mathematical representation of the fully connected layer.

\[
y_i = \text{softmax}(w_i \ast x^i + b_i). \tag{3}
\]

where \(x^i\) represents the input and \(w_i\) and \(b_i\) denote weight and bias factors of the fully connected layer respectively. \(y_i\) is the output feature vector extracted from the CT scan data.

**D. Feature Reduction Using Principal Component Analysis**

Principal component analysis (PCA) is a dimensionality-reduction technique to achieve optimal model performance with reduced dimensions/features of the data [40], [41]. Although this comes at the cost of accuracy but the trick behind PCA is to reduce features with a minimum information loss. Several factors influence performance of a model on imagery data, most notably resource constraints such as RAM, computation power, and storage. To avoid wasting resources, the deep learning algorithm should identify redundant features and remove them from the training set when dealing with large datasets. In literature, researchers have used PCA for feature reduction in various medical applications [42], [43]. Inspired from them, we employed PCA to remove useless features from ICH detection and classification feature set. Distinct features contain a variety of data and discarding them may result in information loss.

Fig. 3 shows the significance of the PCA within the proposed model architecture. The features are referred to as principal components in PCA. As an input, PCA receives a feature vector of size 2048 from the CNN feature extractor. This feature vector is extracted from a single CT slice. If we generate a feature vector for 500 images, then the target feature vector size will be 500 × 2048. This means that when dealing with thousands of CT slices, it will be computationally expensive to handle a bigger feature vector. In this work, we only keep the 90 features with an explained variance of PCA from the 2048 input feature vector and discard the rest. The selected features are then passed to a classifier for ICH subtype prediction.

**E. Extreme Gradient Boosting Classification**

XGBoost is a gradient-based ensemble learning algorithm used for classification over bigger datasets [44]. It is a scalable, supervised learning algorithm based on decision trees. In machine learning, XGBoost has achieved state-of-the-art results for ICH detection [45] and predicting mortality rate of kidney patients [46].

In this study, XGBoost is used to predict the subtype of ICH using the feature vector extracted from CT scan image. The PCA feature selector selects 90 features from each CT image. These features along with a one-hot label are passed as an input to the XGBoost classifier. We performed hyperparameter tuning of the XGBoost classifier. Five ICH classes, the softmax activation function, and the max depth of four is passed to fit the function of the XGBoost classifier. The default values are used for the rest of the parameters.

**F. Hyperparameter Tuning**

Before the training starts, behaviour of the model is determined by the user-supplied hyperparameters such as learning rate, batch size etc. They are considered as external parameters because model cannot change their values during the
FIGURE 6. (a) Shows the samples of original CT image dataset of RNSA-2019. In (b) the generated real CT scan image samples through DCGAN-based data augmentation approach. Our generated CT images have same features as the original dataset.

Automated approaches for building machine learning models have been increasingly prevalent during the past couple of decades. When there are many parameters, it is not possible to manually choose hyperparameters or execute grid searches because these methods rely on human ability and intuition, which can be challenging in particular situations. While you train your knowledge representation system, AutoML, developed by the Google AI research team, automatically learns how different characteristics affect model performance. This alternate approach may perform better with huge datasets.

The hyperparameter configuration for this study is performed manually after more than 200 trial experiments. Table 1 lists down the hyperparameters used for training.
FIGURE 7. Proposed attention-based ResNet-152V2 architecture used for feature extraction.

IV. EXPERIMENTS AND RESULTS

The experiment is carried out using Kaggle Jupyter Notebook and the Python programming language. TensorFlow and Keras libraries are used for implementing deep learning models while Kaggle GPU is leveraged to train and test the deep learning model. The scikit-learn library is used to evaluate the performance of the model.

A. DATASET DESCRIPTION

The comprehensive description of the dataset distribution is provided in Table 2. The dataset images are provided for five independent ICH classes, i.e., EH, SH, IH, IVH, and SAH. The subtypes division of each class in the original RNSA-2019 dataset is extremely imbalanced, and EH subtype only holds 1.5% of the whole dataset. Hence, we generated CT scan images of EH class by employing DCGAN-based data augmentation technique to tackle the class imbalance problem.

We used 434,166 CT scan samples from the balanced version of the RNSA-2019 challenge dataset. The dataset is divided into training, validation, and testing sets. Out of the total samples, 294,551 CT scan images are used for training, 387,81 scans are used for validation, while 100,834 CT scan images are used for testing purposes. All the experiments performed in this work use the same dataset division as mentioned in Table 2.

B. PERFORMANCE EVALUATION METRICS

An accuracy metric is widely used to evaluate the performance of the machine learning and deep learning models. In medical applications, the false positive (FP) and false negative (FN) results are also very critical. We cannot ignore them as the model can misclassify any ICH patient as a healthy person or vice versa. The precision and recall metrics are used to measure FP and FN errors respectively. To measure both precision and recall with equal importance, F1-score is the required metric. F1-score is the harmonic mean of precision and recall. We also calculate true positive rate (TPR) and true negative rate (TNR). These metrics help us in measuring the accuracy of the model with respect to each of the individual ICH classes. The area under the curve (AUC) metric is an indication of how a model is able to correctly predict ICH class. Eq. (4), (5), (6) and (7) provide mathematical formulas...
to calculate the performance evaluation metrics.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4}
\]

\[
\text{TPR (Recall)} = \frac{TP}{TP + FN}, \tag{5}
\]

\[
\text{TNR (Specificity)} = \frac{TN}{TN + FP}, \tag{6}
\]

\[
F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}
\]

where precision and recall are calculated using Eq. (8) and Eq. (5) respectively:

\[
\text{Precision} = \frac{TP}{TP + FP}, \tag{8}
\]

where TP represents true positives cases which are correctly identified as belonging to either of the ICH subtypes. TN stands for true negative cases which are correctly classified as non-ICH patients. In contrast, FP and FN are the misclassified negative and positive ICH patients, respectively.

C. RESULTS AND ANALYSIS

The two sets of results produced in this study are from the DCGAN-based data augmentation model, and the outcome of the proposed model for ICH detection and classification. They are depicted in Fig. 8 and Fig. 9 respectively.

Let us first consider the performance of the DCGAN-based data augmentation module. Fig. 8(a) demonstrates the loss curves of both the generator and discriminator models. The generator generates the fake CT scan images and the discriminator tries to recognise whether the generated image is real or fake. The ultimate goal of the generator is to fool the discriminator and generate such images which resemble the real CT scan images. The blue curve exhibits the training loss for discriminator whereas the green curve represents the training loss for the generator. It can be seen that, with an increase in the number of epochs, a gradual decline in both the discriminator and generator loss is observed. It indicates that the discriminator efficiently distinguishes between real and fake images. Similarly, the generator also manages to fool the discriminator, and generated samples resemble real CT images. This is the typical behaviour of the DCGAN model when used in the supervised learning mode.

Fig. 8(b) shows the accuracy of the data augmentation model during training phase. The blue and the green curves demonstrate the training accuracy of the discriminator in recognising real and fake CT scan samples. The gradual increase in the training accuracy, with the increase in number of epochs, shows how efficiently the generator is trained to generate the CT samples. Hence, discriminator is unable to differentiate the fake CT scan samples generated by the generator from the real ones.

Fig. 9 demonstrates the performance of the proposed attention-based ResNet architecture. Table 3 presents the tabular representation of the data. The results reveal that the proposed model performs well for ICH detection and subtype classification. It achieved superior performance for all evaluation metrics having the best results for the EH subtype. The accuracy for EH class is 99.2%. Moreover, the accuracy for all other ICH subtypes is also greater than 96% which is a satisfactory performance. This achievement is due to the practical configuration and handcrafted attention blocks added to the model. It is an indication that our proposed model effectively discriminates between ICH subtypes due to its robust feature extraction, reduction, and classification mechanism. Similarly, TPR, TNR and AUC are also higher for all the subtypes. EH subtype demonstrates best results for all these metrics having scores of 97.1%, 96.4% and 98.3% respectively. The F1-score is also depicted for all ICH subtypes. The proposed model achieves a maximum F1-score of 96.1% for the EH subtype, while for other ICH types, it is also satisfactory. The higher F1-score reduces the instances of FP and FN and helps to identify the ICH subtypes accurately. The overall results indicate superior performance of the proposed model on the validation dataset.

\begin{table}[h]
\centering
\caption{Proposed model performance of ICH subtypes for different performance metrics.}
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Metrics & EH & IH & IVB & SH & SAH \\
\hline
Accuracy(%) & 99.2 & 97.1 & 96.7 & 96.7 & 96.1 \\
TPR(%) & 97.1 & 94.9 & 95.4 & 96.4 & 95.6 \\
TNR(%) & 96.4 & 95.1 & 92.7 & 91.4 & 90.8 \\
F1-score(%) & 96.1 & 94.3 & 92.7 & 95.5 & 92.1 \\
AUC (%) & 98.3 & 92.9 & 94.1 & 95.0 & 95.3 \\
\hline
\end{tabular}
\end{table}

V. COMPARATIVE ANALYSIS

There are two comparisons conducted in this study. The first is to analyse the impact of the balanced and imbalanced data distributions on the performance of the proposed model. The second is to conduct a comparative analysis of the proposed model with selected benchmark models.

A. PERFORMANCE EVALUATION OF THE PROPOSED MODEL FOR BALANCED AND IMBALANCED DATA DISTRIBUTIONS

The performance of the proposed model for balanced and imbalanced dataset is given in Table 4. The results show that the overall performance of the proposed model is better for the balanced dataset than imbalanced dataset. The reported accuracy score for balanced dataset is 97.2% whereas it is 94.3% for the imbalanced dataset. Similarly, the model obtains better values for TPR and AUC metrics for the balanced data distribution. They are 95.9% and 95.1% for balanced dataset, which were 94.4% and 90.3% for the imbalanced dataset respectively. The F1-score, which combines both precision and recall is 94.1% for the balanced dataset and 87.7% for the imbalanced dataset. The TNR metric of 93.2% is the only exception where its value for imbalanced data is better than the balanced data.

The better performance of the proposed model over balanced dataset highlights the importance of the balanced data distribution for the ICH subtype classification. The imbalanced dataset contains skewed data distribution, which raises...
overfitting issues. Most of the studies in the literature display poor performance because they have used imbalanced datasets for experimentation. Moreover, our DCGAN-based data augmentation model uses binary cross-entropy loss function which performs better in terms of convergence rate and provides stable learning.

B. COMPARATIVE ANALYSIS OF THE PROPOSED MODEL WITH BENCHMARK TECHNIQUES

The four well-known benchmark models are chosen for the comparative analysis. It also includes state-of-the-art model that acts as a baseline for analysis in this study. The following section provides description of these models.

1) BENCHMARK MODELS

a: Double-Branch ResNet-50 and RF

This model uses a double-branch-CNN with a pre-trained ResNet-50 architecture for feature extraction [10]. The extracted features with ground truth are passed to support

TABLE 4. Comparative analysis of proposed model for balanced and imbalanced data distributions.

| Performance metrics | Balanced dataset | Imbalance dataset |
|---------------------|------------------|------------------|
| Accuracy (%)        | 97.3             | 94.3             |
| TPR (%)             | 95.9             | 96.4             |
| TNR (%)             | 93.2             | 95.3             |
| F1-score (%)        | 92.1             | 92.7             |
| AUC (%)             | 95.3             | 90.3             |

FIGURE 8. Training loss and accuracy of the DCGAN-based data augmentation model.

FIGURE 9. Performance evaluation of the proposed model for ICH classification.
vector machine (SVM) and random forest (RF) classifiers for class prediction. The RF classifier displayed better results as compared to the SVM for ICH subtype classification.

b: MOBILENETV2

It is one of the best pre-trained models for Keras applications trained on the ImageNet dataset. MobileNetV2 has 3.5M parameters and a depth of 105 layers. It can classify 1000 classes with a top-5 accuracy of 90.01% on ImageNet validation dataset [47]. It works incredibly well for computer vision applications such as image classification, feature extraction, and segmentation etc. The model shows efficient results in image classification because it employs inverted residual blocks with fewer parameters. We used transfer learning approach to fine-tune it for ICH subtype classification.

c: XCEPTION

It is an 81-layer deep CNN network and the top model of the Keras application trained on the ImageNet dataset [48]. Xception has 94.5% of the top-5 accuracies on ImageNet validation dataset. Due to the use of pointwise and depth-wise convolutional, it is a state-of-the-art model for object detection and feature extraction applications. Using pre-trained weights from ImageNet, we fine-tuned the Xception model to identify the ICH subtypes.

d: RESNEXT-101 AND BILSTM NETWORK

It is a state-of-the-art deep learning model used for ICH detection and subtype classification. It comprises a hybrid CNN and bidirectional long short-term memory (BiLSTM) network [12]. The proposed model employs the pre-trained ResNeXt-101 architecture and BiLSTM for prediction in 3D Scans at the slice level and scan level. The model also uses Grad-Cam visualizations to better diagnose and cross-validate the performance of the model.

2) DISCUSSION

Table 5 presents the comparison between the epidural hemorrhage (EH) class results of the proposed model and the benchmark techniques for different performance metrics. The comparison with other subtypes is illustrated in Fig. 10. The results (bold and underlined) show an excellent achievement of our proposed model as compared to benchmark techniques. It achieves an accuracy of 99.2%, which is the best performance among all benchmark studies. In contrast, ResNet-50 and RF model obtains the lowest accuracy of 74.3% due to redundant feature extraction, and imbalanced training dataset. It also suffers from the curse of dimensionality during fine-tuning on the CT image dataset. The benchmark model MobileNetV2 achieves an accuracy of 86.9% for ICH detection. It uses the inverted residual blocks and skip connections introduced for an efficient lightweight mobile vision application. A thin layer with a modified residual block in MobileNetV2 helps to make the network deeper. The Xception model uses separable convolutional blocks with no non-linearity. It is the top model from Keras application for multiclass image classification on benchmark datasets with pre-trained weights. It achieves an accuracy of 88.0% for ICH classification, which denotes a better model performance. The performance of our proposed model for other performance metrics is also satisfactory.

The accuracy comparison of the proposed model for ICH subtype classification with the benchmark schemes is
TABLE 5. Comparative analysis of the proposed model with benchmark techniques for EH class.

| Model                              | Accuracy (%) | TPR (%) | TNR (%) | F1-score (%) |
|------------------------------------|--------------|---------|---------|--------------|
| ResNet-50 and RF                   | 76.3         | 68.6    | 79.9    | 72.7         |
| MobileNetV2                        | 86.9         | 93.3    | 87.6    | 87.2         |
| Xception                           | 85.0         | 87.3    | 88.8    | 87.8         |
| ResNeXt-101 and BiLSTM             | 96.4         | 94.8    | 81.2    | 90.8         |
| Proposed CNN                       | **99.3**     | **97.1**| **96.4**| **96.1**     |

FIGURE 11. F1-score comparison of the proposed model with benchmark techniques for EH class.

depicted in Fig. 10. The proposed model obtains the best accuracy as compared to the existing models for all ICH subtypes. Its best performance is for the EH class having an accuracy score of 99.2%, while its worst performance is for the SH subtype, which is a 96.1% accuracy score. However, both these results are best among corresponding subtypes in all the benchmark techniques. This performance highlights that the proposed model accurately differentiates between acute ICH subtypes due to its attention-based hybrid feature extraction mechanism. In addition, the PCA feature reduction technique (which captures only pertinent features from CT images) and the XGBoost classifier also enhances the classification of ICH subtypes. Consequently, the proposed hybrid scheme efficiently performs the ICH detection and subtype classification. The comparison of F1-scores from the proposed and benchmark models is illustrated in Fig. 11. The F1-score combines both precision and recall in a single metric. The proposed model achieves a 96.1% F1-score, which is best among all the benchmark models. It signifies that the proposed model is appropriate for use in real-world medical applications to accurately identify the ICH and its subtypes.

Among all the comparisons drawn above, the state-of-the-art ResNeXt-101 and BiLSTM model displayed best performance. Hence, it is selected as a baseline model for comparison. When compared with the proposed model for the EH class performance, our model beats it in all metrics as shown in the Table 5. The accuracy of our model for other ICH subtypes is also better than the baseline model as seen in Fig. 10. Although we managed to beat the baseline model but our results are not superior by a wide margin, especially for accuracy and TPR. The competitive performance of the baseline model is due to adapting hybrid learning procedure. It utilizes the strengths of both ResNeXt-101 and BiLSTM algorithms for feature extraction, which makes it a better model among its competitor models.

The proposed model has several limitations. This work only considers a single dataset for experimentation, results may vary if it is tested on a different dataset. It only tests a DCGAN-based model for data augmentation, other variants of GAN may produce even better results. Moreover, our model is data-hungry, it needs large amount of data to train.

VI. CONCLUSION AND FUTURE WORK

In this paper, an attention-based deep learning model is designed for the detection and subtype classification of acute intracranial hemorrhage (ICH) in CT scans. The proposed model combines attention mechanism with ResNet-152V2 architecture for feature extraction, PCA for feature selection and XGBoost for classification. By leveraging DCGAN to produce realistic CT images, the problem of the significant lack of epidural class ICH samples is resolved. The study is undertaken using the RNSA-2019 dataset, which was made accessible by the Radiological Society of North America (RSNA). The proposed model outperforms benchmark models such as ResNet-50 and RF, MobileNetV2, Xception, ResNeXt-101, and BiLSTM in terms of accuracy and F1-score. The findings reveal that our model achieves accuracies of 99.2%, 97.1%, 96.7%, 96.7% and 96.1%, for detecting epidural hemorrhage (EH), intraparenchymal hemorrhage (IH), intraventricular hemorrhage (IVH), subdural hemorrhage (SH), and subarachnoid hemorrhage (SAH) subtypes respectively. Moreover, F1-score of 96.1% for EH class is also best among all benchmarks models. The superior performance of the proposed model makes it appropriate for use in real-world applications.

To further enhance the detection results, we will examine the performance of our proposed scheme using automated hyperparameter tuning techniques in the future. In addition, we will use state-of-the-art vision transformer (ViT) model to test the precision of ICH multi-class classification. Different GAN algorithms can be used to balance the RNSA-2019 dataset, and ICH CT data from nearby hospitals can also be taken into account.

REFERENCES

[1] M. Alja’Afreh, “A QoE model for digital twin systems in the era of the tactile internet,” Ph.D. dissertation, School Electr. Eng. Comput. Sci., Univ. Ottawa, Ottawa, ON, Canada, 2021.
[2] R. Ferdousi, F. Laamarti, M. A. Hossain, C. Yang, and A. E. Saddik, “Digital twins for well-being: An overview,” Digit. Twin, vol. 1, no. 7, p. 7, Oct. 2021.
[3] A. E. Saddik, F. Laamarti, and M. Alja’Afreh, “The potential of digital twins,” IEEE Instrum. Meas. Mag., vol. 24, no. 3, pp. 36–41, May 2021.
[4] H. Ye, F. Gao, Y. Yin, D. Guo, P. Zhao, Y. Lu, X. Wang, J. Bai, K. Cao, Q. Song, H. Zhang, W. Chen, X. Guo, and J. Xia, “Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network,” Eur. Radiol., vol. 29, no. 11, pp. 6191–6201, Nov. 2019.
[5] X. Wang, T. Shen, S. Yang, J. Lan, Y. Xu, M. Wang, J. Zhang, and X. Han, "A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans," *NeuroImage, Clin.*, vol. 32, Jan. 2021, Art. no. 102785.

[6] R. A. Rava, S. E. Seymour, M. E. LaQue, B. A. Peterson, K. V. Snyder, M. Mokin, M. Waqas, Y. Hoi, J. M. Davies, E. I. Levy, A. H. Siddiqui, and C. N. Ionita, "Assessment of an artificial intelligence algorithm for detection of intracranial hemorrhage," *World Neurosurg.*, vol. 150, pp. e209–e217, Jun. 2021.

[7] A. M. Dawud, K. Yurtkan, and H. Oztoprak, "Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–12, Jun. 2019.

[8] P. Gruschwitz, J.-P. Grunz, and T. Van Hoa, "Brain hemorrhage diagnosis by using deep learning," in *Proc. 40th Annu. Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 1–7. Apr. 2019.

[9] A. Sage and P. Badura, "Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest," *Appl. Sci.*, vol. 10, no. 21, p. 7577, Oct. 2020.

[10] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, and L. Wang, "Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans," *Sensors*, vol. 19, no. 9, p. 2167, May 2019.

[11] M. Burduja, R. T. Ionescu, and N. Verga, "Accurate and efficient intracranial hemorrhage detection and subtype classification in 3D CT scans with convolutional and long short-term memory neural networks," *Sensors*, vol. 20, no. 19, p. 5611, Oct. 2020.

[12] M. Togacar, Z. Comert, B. Ergen, and U. Budak, "Brain hemorrhage detection based on heat maps, autoencoder and CNN architecture," in *Proc. int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2019, pp. 1–7.

[13] X. Ma, Y. Bai, Y. Lin, X. Hong, T. Liu, L. Ma, M. E. Haakec, J. Zhou, J. Wang, and M. Wang, "Amide proton transfer magnetic resonance imaging in detecting intracranial hemorrhage at different stages: A comparative study with susceptibility weighted imaging," *Sci. Rep.*, vol. 7, no. 1, pp. 1–7, May 2017.

[14] T. D. Phong, H. N. Duong, H. T. Nguyen, N. T. Trong, V. H. Nguyen, T. Van Hou, and V. N. Stasel, "Brain hemorrhage diagnosis by using deep learning," in *Proc. int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2019, pp. 1–7.

[15] J. Ljungqvist, S. Candeford, M. Persson, L. Jönsson, T. Skoglund, and M. Elam, "Clinical evaluation of a microwave-based device for detection of traumatic intracranial hemorrhage," *J. Neurotrauma*, vol. 34, no. 13, pp. 2176–2182, 2017.

[16] J. Muschelli, E. M. Sweeney, N. L. Ullman, P. Vespa, D. F. Hanley, and C. M. Graimiczak, "PicHPERF-3C: Primary intracranial hemorrhage probability estimation using random forests on CT," *NeuroImage, Clin.*, vol. 14, pp. 379–390, Jan. 2017.

[17] L. Xu, X. Tao, W. Liu, Y. Li, J. Ma, T. Lu, B. Han, B. Liu, Y. Zhao, J. Li, and J. Zhao, "Portable near-infrared rapid detection of intracranial hemorrhage in Chinese population," *J. Clin. Neurosci.*, vol. 40, pp. 136–146, Jun. 2017.

[18] M. R. Arbabsirani, B. K. Fornwalt, G. M. Mongelluzzo, J. D. Suarez, B. E. Leise, A. P. Richman, and G. I. M. Oprea, "A machine learning based system for real-time detection and classification of subdural hematomas from head CT images," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 450–461, Jun. 2019.

[19] K. Jnawali, M. R. Arbabsirani, A. E. Ullaho, N. Rao, and A. A. Patel, "Automatic classification of radiological report for intracranial hemorrhage," in *Proc. IEEE 13th Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2019, pp. 187–190.

[20] N. T. Nguyen, D. Q. Tran, N. T. Nguyen, and H. Q. Nguyen, "A CNN-LSTM architecture for detection of intracranial hemorrhage on CT scans," *medRxiv*, vol. 2020, pp. 1–4, Jan. 2020.

[21] H. Chen, S. Khan, B. Kour, S. Nazir, W. Liu, and A. Hussain, "A smart machine learning model for the detection of brain hemorrhage diagnosis based on Things of smart cities," *Complexity*, vol. 2020, pp. 1–10, Sep. 2020.

[22] D. T. Gint, "Analysis of head CT scans flagged by deep learning software for acute intracranial hemorrhage," *NeuroImage, Clin.*, vol. 62, no. 3, pp. 335–340, Mar. 2020.

[23] J. He, "Automated detection of intracranial hemorrhage on head computed tomography with deep learning," in *Proc. 10th Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2019, pp. 121–127.

[24] J. L. Meek, T. Vicar, and R. Jakubick, "Weakly supervised deep learning-based intracranial hemorrhage localization," 2021, arXiv:2105.00781.

[25] P. Singh, V. Khanna, and M. Kamal, "Hemorrhage segmentation by fuzzy c-mean with modified level set on CT imaging," in *Proc. 5th Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2019, pp. 550–555.

[26] R. F. Mansour and N. O. Aljehane, "An optimal segmentation with deep learning based inception network model for intracranial hemorrhage diagnosis," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13831–13843, Oct. 2021.

[27] B. Rao, V. Zohrabian, P. Cedeno, A. Saha, J. Pahade, and M. A. Davis, "Utility of artificial intelligence tool as a prospective radiology peer reviewer—Detection of unreported intracranial hemorrhage," *Academic Radiol.*, vol. 28, no. 1, pp. 85–93, Jan. 2021.

[28] Y. Wu, M. P. Supanic, and J. Deng, "Ensemble deep neural network for intracranial hemorrhage detection and subtype classification on noncontrast CT images," *J. Artif. Intell. Med. Sci.*, vol. 2, nos. 1–2, pp. 12–20, 2021.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 134–144, 2020.

[30] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, arXiv:1511.06434.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), 2016, pp. 770–778.

[32] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Statistic*, vol. 2, no. 4, pp. 433–459, Jul/Aug. 2010.

[33] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent development," *Philos. Trans. Roy. Soc. A Math., Phys. Eng. Sci.*, vol. 374, no. 2065, 2016, Art. no. 20150202.

[34] G. M. Ibrahim, B. R. Morgan, and R. L. Macdonald, "Patient phenotypes associated with outcomes after aneurysmal subarachnoid hemorrhage: A principal component analysis," *Stroke*, vol. 45, no. 3, pp. 670–676, Mar. 2014.

[35] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, pp. 274, Feb. 2020.

[36] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (KDD), Aug. 2016, pp. 785–794.
A. Hussain et al.: Attention-Based ResNet Architecture for Acute Hemorrhage Detection and Classification

[45] D. Abe, M. Inaji, T. Hase, S. Takahashi, R. Sakai, F. Ayabe, Y. Tanaka, Y. Otomo, and T. Machara, “A prehospital triage system to detect traumatic intracranial hemorrhage using machine learning algorithms,” *JAMA Netw. Open*, vol. 5, no. 6, Jun. 2022, Art. no. e2216393.

[46] J. Liu, J. Wu, S. Liu, M. Li, K. Hu, and K. Li, “Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model,” *PLoS ONE*, vol. 16, no. 2, Feb. 2021, Art. no. e0246306.

[47] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[48] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

**AFTAB HUSSAIN** received the B.S. degree in computer science from The Islamia University of Bahawalpur, Pakistan, and the M.S. degree in computer science from COMSATS University Islamabad, Islamabad Campus, in 2022. He is currently working as a Visiting Faculty in computer science and IT at Emerson University Multan, Pakistan. His primary research interests include machine learning, deep learning, medical imaging, computer vision, natural language processing, artificial intelligence in smart cities, medical healthcare, and the IoT.

**MUHAMMAD USMAN YASEEN** is currently working as an Assistant Professor at COMSATS University Islamabad (CUI), Islamabad, Pakistan. His current research interests include video analytics, big data analysis, machine learning, and distributed systems.

**MUHAMMAD IMRAN** received the Graduate degree in software engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2006, and the master’s degree in software engineering and the Ph.D. degree in computer science from the University of Southampton, U.K., in 2009 and 2015, respectively. He was a Lecturer with COMSATS University Islamabad (CUI), Islamabad, Pakistan, from 2007 to 2008, where he currently is an Assistant Professor with the Department of Computer Science. His research interests include artificial intelligence, machine learning, natural language processing, and semantic web.

**MUHAMMAD WAQAR** received the B.S. and M.S. degrees in computer engineering from COMSATS University Islamabad (CUI), Pakistan, in 2011 and 2015, respectively, and the Ph.D. degree from Sejong University, Seoul, South Korea, in 2020. From 2012 to 2017, he worked as a Team Leader and a Site Administrator on the Grid Computing Project of CERN (European Organization for Nuclear Research) at CUI, where he is currently working as an Assistant Professor at the Department of Computer Science. His research interests include 5G cloud systems, information security, machine learning, and deep learning.

**ADNAN AKHUNZADA** (Senior Member, IEEE) is currently a Professional Member of ACM, with 15 years of research and development experience both in ICT industry and academia. He has a proven track record of high impact published research (i.e., U.S. patents, journals, IEEE TRANSACTIONS, reputable magazines, book chapters, conferences, and conference proceedings) and commercial products. His experience as an Educator and a Researcher is diverse that includes work as a Lecturer, a Senior Lecturer, a Year Tutor, and an Occasional Lecturer at other engineering departments, as an Assistant Professor at COMSATS University Islamabad (CUI), a Senior Researcher at RISE SICS Västerås AB, Sweden, as a Research Fellow and the Scientific Leader of DTU Compute, Technical University of Denmark (DTU), as the Course Director of Ethical Hacking at The Knowledge Hub Universities (TKH), Coventry University, U.K., an Associate Professor at USM, and a Visiting Professor having mentorship of graduate students and supervision of academic and research and development projects both at UG and PG levels. His research interests include cyber security, secure future internet, artificial intelligence (i.e., machine learning, deep learning, and reinforcement learning), large scale distributed systems (i.e., edge, fog, cloud, and SDN), the IoT, the IIoT, QoS/QoE, affective computing, industry 4.0, and the Internet of Everything (IoE). He is also a member of technical program committee of varied reputable conferences, journals, and editorial boards.

**MOHAMMAD AL-JAAFREH** received the master’s and Ph.D. degrees from the University of Ottawa. He is currently working as a Research Scientist at the Mohamed bin Zayed University of Artificial Intelligence and a Research and Development Associate at the Cybersecurity Division, Department of Natiation Defence, Royal Military College of Canada. He is also an Adjunct Professor at the Faculty of Engineering, University of Ottawa, ON, Canada. During his Graduate studies, he received some prestigious scholarships, such as the International Germany Jordanian University Graduate Sponsorship and Ontario Graduate Scholarship. His research interests include modern multimedia systems and communication, AI and cybersecurity for intelligent things, and QoE for Internet of Skills (IoS) based on affective computing.

**ABDULMOTAUDES EL SADDIK** is currently the Acting Chair of the Mohamed bin Zayed University of Artificial Intelligence and a Distinguished Professor with the School of Electrical Engineering and Computer Science, University of Ottawa. He is also an internationally recognized scholar who has made strong contributions to the knowledge and understanding of multimedia computing, communications, and applications, particularly in the area of digital twins.

He is a fellow of the Royal Society of Canada, the Canadian Academy of Engineering, and the Engineering Institute of Canada. He is an ACM Distinguished Scientist and has received several awards, including the Friedrich Wilhelm Bessel Award from the German Humboldt Foundation and the IEEE Instrumentation and Measurement Society Technical Achievement Award. He also received IEEE Canada C. C. Gotlieb (Computer) Medal and A. G. L. McNaughton Gold Medal for important contributions to the field of computer engineering and science and the 2021 IEEE TCSC Achievement Award for Excellence in Scalable Computing.