

Article

Machine Learning Model for Intracranial Hemorrhage Diagnosis and Classification

Sundar Santhoshkumar ¹, Vijayakumar Varadarajan ^{2,*}, S. Gavaskar ³, J. Jegathesh Amalraj ⁴ and A. Sumathi ⁵

¹ Department of Computer Science, Alagappa University, Karaikudi 630003, Tamil Nadu, India; santhoshkumars@alagappauniversity.ac.in

² School of Computing Science and Engineering, The University of New South Wales, Sydney, NSW 2025, Australia

³ Department of Computer Applications, Bharathiar University, Coimbatore 641046, Tamil Nadu, India; gavas_blue@yahoo.co.in

⁴ Department of Computer Science, Government Arts and Science College, Cuddalore, Tittagudi 606106, Tamil Nadu, India; amal.jas@gmail.com

⁵ Department of Computer Science and Engineering, SRC, SASTRA Deemed to be University, Kumbakonam 612001, Tamil Nadu, India; sumathi@src.sastra.edu

* Correspondence: vijayakumarvaradarajan@gmail.com

Abstract: Intracranial hemorrhage (ICH) is a pathological disorder that necessitates quick diagnosis and decision making. Computed tomography (CT) is a precise and highly reliable diagnosis model to detect hemorrhages. Automated detection of ICH from CT scans with a computer-aided diagnosis (CAD) model is useful to detect and classify the different grades of ICH. Because of the latest advancement of deep learning (DL) models on image processing applications, several medical imaging techniques utilize it. This study develops a new densely connected convolutional network (DenseNet) with extreme learning machine (ELM) for ICH diagnosis and classification, called DN-ELM. The presented DL-ELM model utilizes Tsallis entropy with a grasshopper optimization algorithm (GOA), named TEGO, for image segmentation and DenseNet for feature extraction. Finally, an extreme learning machine (ELM) is exploited for image classification purposes. To examine the effective classification outcome of the proposed method, a wide range of experiments were performed, and the results are determined using several performance measures. The simulation results ensured that the DL-ELM model has reached a proficient diagnostic performance with the maximum accuracy of 96.34%.

Keywords: multilevel thresholding; DenseNet; deep learning; ICH diagnosis; CT images; computer-aided diagnosis



check for updates

Citation: Santhoshkumar, S.; Varadarajan, V.; Gavaskar, S.; Amalraj, J.J.; Sumathi, A. Machine Learning Model for Intracranial Hemorrhage Diagnosis and Classification.

Electronics **2021**, *10*, 2574. <https://doi.org/10.3390/electronics10212574>

Academic Editor: Bhanu Prakash K.N.

Received: 16 September 2021

Accepted: 18 October 2021

Published: 21 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intracranial hemorrhage (ICH) is an important and a severe disease that paves the way for heart disease and stroke. ICH mostly affects severely overweight people and the mortality rate has been enhanced progressively within a limited time period. Moreover, it occurs in multiple intracranial blocks, which are caused by many external factors. In order to treat ICH, a neuro-imaging mechanism is available for examining the position and quantity of hemorrhage and its impending cerebral damage, which helps inpatient treatment [1]. However, it is externally affected in the brain parenchyma (extra-axial). This intra-axial as well as extra-axial hemorrhage is impossible to treat unless it is discovered in the earlier stage. For instance, intra-axial hemorrhage influences severely overweight people in the United States with maximum fatality [2]. Clinical admission of ICH has been enhanced drastically because of the growing population, expensive lifestyle, and abnormal blood pressure management. Furthermore, later diagnosis of ICH cause one of the serious health effects that results in massive death within a short period of time, Computed tomography (CT) screening is the general mechanism used for diagnosing ICH

accurately, as well as diagnosing it early, which relies on the robustness of CT with respect to quick interpretation of ICH.

The interpretation of radiological work depends upon the preference for any patient to be tested as either an inpatient or outpatient. Typically, Stat works are interpreted with limited time, whereas regular outpatient examination takes the maximum duration, which relies on an accessible radiology system. ICH occurs in an outpatient setting, albeit with low frequency when compared with an inpatient or emergency department setting. For instance, an aged outpatient on anticoagulation remedy suffers from the risk of ICH [3–5]. Interestingly, the primary signs might be vague, which prompts non-emergent, regular head CT. Furthermore, CT is defined as a popular non-invasive and efficient imaging technique for ICH prediction. Hemorrhage is examined in non-contrast CT as blood has a high density (Hounsfield unit, HU) when compared with alternate brain cells; however, it is limited for bones. An exact analysis of bleeding is very serious for medical interventions. Moreover, the estimation of head CT is required for patients admitted in emergencies. Simultaneously, a primary interpretation of head CT is done by junior radiologists and trainee radiologists for emergency patients. Finally, initial interpretations are made by expert radiologists.

An automated triage of imaging studies, which apply computer models capable of predicting ICH with enhanced results, was used. A quality enhancement tool was employed for automated management for early interpretation of imaging works with supposed ICH as well as optimization of the radiology task. Computer vision and machine learning (ML) methodologies are suitable for learning and predicting the patterns. Specifically, the DL algorithm is a kind of ML model that has been leveraged for automatic classification operations like natural language processing (NLP), audio analysis, and object prediction [6–8]. Progressive development in “augmented” diagnostic vision with ML in the clinical field. For instance, DL models are used for diagnosing diabetic retinopathy (DR) from retinal images, breast cancer from mammograms, and so on. The published applications involve the prediction and diagnosis of skin cancer, pulmonary lumps, and cerebral micro-HM. In spite of the studies demonstrating the efficiency of ML for diagnostic medicine and radiology, medical implementation of DL technology remains unexplored [9,10].

Automated identification of ICH using CT scans using computer-aided diagnosis (CAD) models can be employed to increase the detection rate in a short period of time. As the quantity of neuro imaging data obtainable for the design of these solutions is normally restricted, this paper designs an effective densely connected convolutional network (DenseNet) with an extreme learning machine (ELM) for ICH classification and diagnoses, called DL-ELM. The presented method comprises several sub processes, namely, classification, pre-processing, segmentation, feature extraction, and so on. The DL-ELM model undergoes a pre-processing step, where the input data from the NIfTI files are transformed into JPEG format. Next, Tsallis entropy with a grasshopper optimization algorithm (GOA), named TEGO, is used for image segmentation. Afterward, the DenseNet algorithm is applied to identify the useful set of feature vectors and, finally, the ELM is employed to categorize the ICH into different class labels. A detailed analysis of the experimental results takes place to determine the performance of the DL-ELM technique.

2. State-of-the-Art Approaches to ICH Diagnosis

Many traditional and DL algorithms are explained in this work. According to the traditional ML models, Yuh et al. [11] developed a threshold-reliant methodology for the prediction of ICH. A technique predicted an ICH sub-types, which depend upon the position, structure, and volume. Developers have optimized the threshold value under the application of retrospective instances of CT scans and determined on CT scans of subjects with traumatic brain injury (TBI). Consequently, maximum sensitivity and specificity are accomplished for ICH prediction and intermediate accuracy is accomplished while predicting ICH sub-types. Alternatively, Li et al. [12] projected two models to segment subarachnoid hemorrhage (SAH) space and applied segmented regions for the purpose of

forecasting SAH. In this approach, CT scans are employed to train and test mechanisms. Effective performance was addressed with the help of the Bayesian decision model with testing SE, SP, and accuracy.

Based on the DL models, convolutional neural networks (CNNs) and corresponding variants are deployed in [13], which depend upon the fully convolutional networks (FCNs) approach. Here, spatial dependence among adjacent slices was assumed under the application of random forest (RF) or recurrent neural network (RNNs). Moreover, developers have applied an extended version of CNNs to compute a complete CT scan or interpolation layer. Alternate technologies are one-stage, which means that it does not apply spatial dependency among the slices. Prevedello et al. [14] projected two methodologies related to CNNs. The primary approach concentrated on ICH prediction, hydrocephalus, and mass effect under the scan level, whereas alternate models are established for predicting malicious acute infarcts.

Chilamkurthy et al. [15] projected four models for forecasting ICH subtypes, namely, midline shift, mass effect, and calvarial fractures. They validated and trained their processes on a massive dataset with maximum CT scans, correspondingly. The two datasets are utilized mainly for testing, where one model has partial scans that are available in a common dataset named as CQ500.

Clinical radiology reports are employed as the gold-standard for labelling of the trained CT scans and authentication of CT scans. This medical report is employed for scanning and the NLP model is also employed for testing the scanning reports and annotated using massive votes of ICH subtypes addressed by three specialized radiotherapists. Diverse deep methods were employed for four predictive types, namely, ResNet18 undergoes training with five parallel FC layers as the output layer. Ye et al. [16] designed a 3D joint convolution and recurrent neural network (CNN-RNN) for the purpose of classifying and predicting ICH. Hence, the entire structure of this technique is the same as in the method developed by Grewal et al. [17]. VGG-16 was applied as a CNN mechanism and bi-GRU was applied as an RNN model. Therefore, the RNN layer performs a similar function to the slice interpolation approach, as presented by Lee et al. [18], although it is effective in terms of adjacent slices applied in classifications. Hence, it is trained and verified on CT scans and sampled. A more precise slice-level was attained in ICH detection.

In line with this, Jnawalia et al. [19] applied a TL method on an ensemble of four popular CNN methodologies for forecasting ICH sub-types and bleeding points. A spatial dependency from adjacent slices is regarded as a slice interpolation framework. This ensemble model undergoes training and is verified under the application of a dataset using CT scans, and is tested through a retrospective database using CT scans as well as a prospective dataset. As a result, ICH prediction is used to develop a better area under the ROC curve (AUC), specificity, and sensitivity. Thus, the newly developed approach resulted in minimum SE for classifying ICH sub-types.

3. Proposed Methodology

In this study, a new DL-ELM model is introduced for the diagnosis and classification of ICH. Initially, the input data from the NIFTI file are transformed into JPEG images. The pre-processed data are segmented using the TEGOA model, and then features are extracted using the DenseNet model. Finally, the ELM method is employed for classifying the different class labels of ICH. The working principle is exhibited in Figure 1 and the algorithms are discussed in the following subsections.

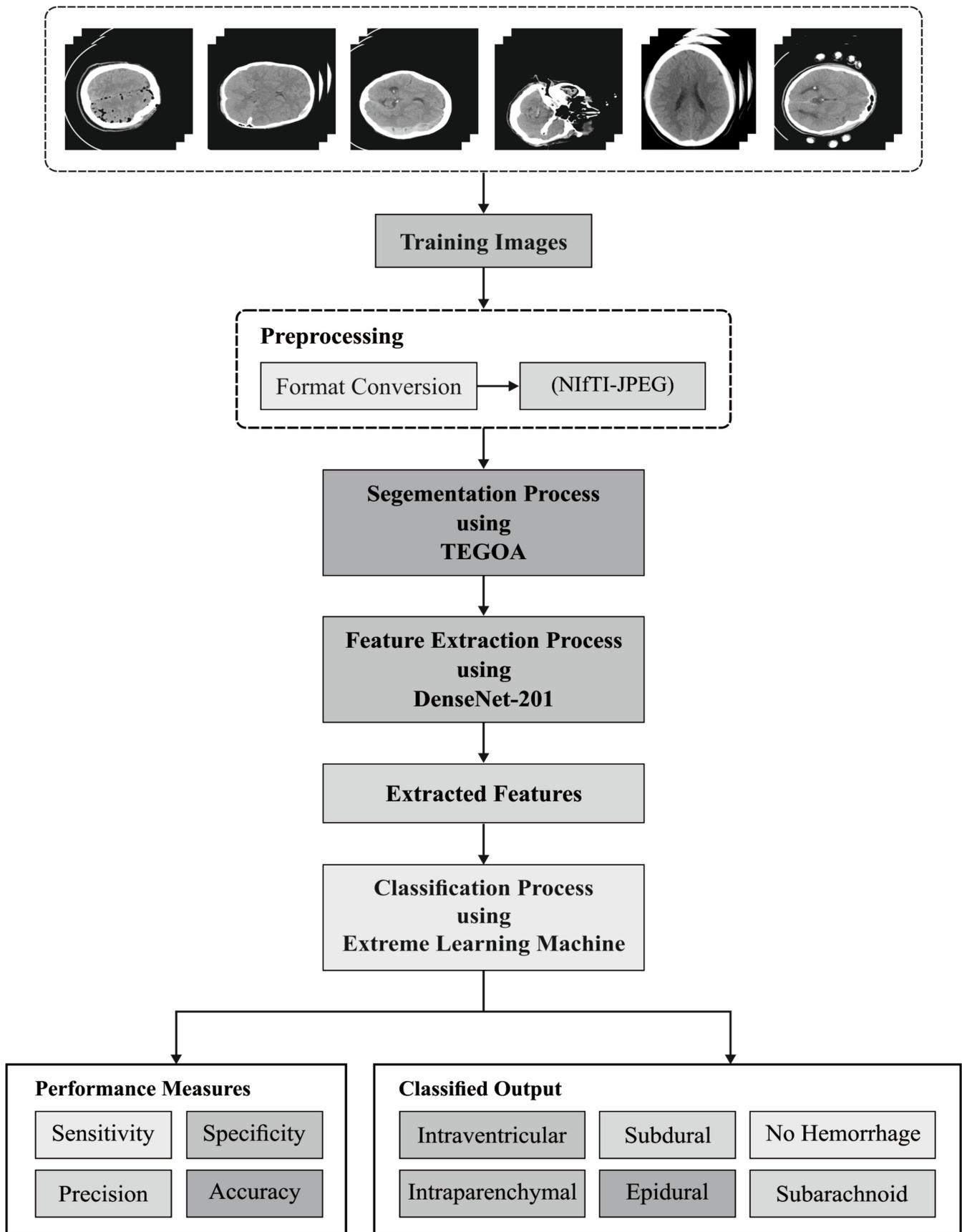


Figure 1. Overall process of the DL-ELM (advancement of deep learning model based extreme learning machine).

3.1. TEGOA-Based Segmentation Process

Primarily, the input data are preprocessed and then the segmentation process is carried out. Entropy is relevant to the chaos value inside a network. Initially, Shannon applied entropy is used to measure the uncertainties of the data involved in a system. It is recommended that, after a physical system is divided into statistically free A & B subsystems, an entropy measure is determined as follows:

$$S(A + B) = S(A) + S(B) \tag{1}$$

According to Shannon’s strategy, a non-extensive entropy paradigm has been presented by Tsallis and is expressed as follows:

$$S_q = \frac{1 - \sum_{i=1}^T (p_i)^q}{q - 1} \tag{2}$$

where T denotes the system’s capability, q implies the entropic index, and p_i refers to the possibility of all states i . In general, Tsallis entropy S_q satisfies Shannon’s entropy if $q \rightarrow 1$. An entropy score is defined as a pseudo additive rule, as given below:

$$S_q(A + B) = S_q(A) + S_q(B) + (1 - q) \cdot S_q(A) \cdot S_q(B) \tag{3}$$

The Tsallis entropy is assumed for identifying effective thresholds of an image [20]. Assume an image with L gray level from $\{0, 1, \dots, L - 1\}$ with likelihood distribution $p_i = p_0, p_1, \dots, p_{L-1}$. Therefore, Tsallis multilevel thresholding is attained by applying the given objective function:

$$f(T) = [t_1, t_2, \dots, t_{e-1}] = \text{argmax} \left[S_q^A(T) + S_q^B(T) + \dots + S_q^K(T) + (1 - q) \cdot S_q^A(T) \cdot S_q^B(T) \dots S_q^K(T) \right] \tag{4}$$

where

$$S_q^A(T) = \frac{1 - \sum_{i=0}^{t_1-1} \left(\frac{P_i}{P^A} \right)^q}{q - 1}, P^A = \sum_{i=0}^{t_1-1} P_i \tag{5}$$

$$S_q^B(T) = \frac{1 - \sum_{i=t_1}^{t_2-1} \left(\frac{P_i}{P^B} \right)^q}{q - 1}, P^B = \sum_{i=t_1}^{t_2-1} P_i \tag{6}$$

$$S_q^K(T) = \frac{1 - \sum_{i=t_{k-1}}^{L-1} \left(\frac{P_i}{P^K} \right)^q}{q - 1}, P^K = i = t, \sum_{i=t_{k-1}}^{L-1} P_i \tag{7}$$

In the case of the multi-level thresholding model, it has to compute an optimal threshold value T that enhances an objective function $f(T)$. In this case, $(f(T))$ maximization has been performed under the application of GOA.

The GOA accelerates similar to the normalized swarm nature of grasshoppers. Likewise, in swarm methods, a grasshopper implies a candidate solution that is generated randomly at the time of initialization; furthermore, using the evaluation function, an optimal grasshopper would be considered as a leader. A leader attracts neighbouring grasshoppers towards it. X_i implies the place of the i th grasshopper in n dimension space. The numerical formation of GOA is depicted as follows:

$$X_i = S_i + G_i + A_i, \tag{8}$$

where S_i represents a social communication as described in Equation (9), G_i depicts a gravity force illustrated in Equation (11), and A_i denotes a wind advection demonstrated in Equation (12).

In the case of grasshopper action, social communication S_i plays an important role, which is attained from Equation (9).

$$S_i = \sum_{j=1, j \neq i}^N s(d_{ij}) \hat{d}_{ij}, \tag{9}$$

where $\hat{d}_{ij} = \frac{x_j - x_i}{d_{ij}}$ denotes a unit vector of two grasshoppers, d_{ij} refers to the Euclidean distance among two grasshoppers, s is accomplished using $d_{ij} = |x_j - x_i|$, and s signifies a function for estimating the intensity of social communication and is evaluated as follows:

$$s(r) = fe^{-\frac{r}{l}} - e^{-r}, \tag{10}$$

where f signifies the intensity of attraction and l refers to an attractive length scale. A study on the nature of grasshoppers with diverse measures of l and f also identifies that the distance between grasshoppers within $[0, 2.079]$ can be repulsive. It becomes a comfort zone. The function used for determining the gravity factor is represented as follows:

$$G_i = -ge_g^\Lambda, \tag{11}$$

where g denotes a gravitational constant and e_g^Λ implies a unit vector. The estimated equation of wind direction is formulated as follows:

$$A_i = ue_w^\Lambda, \tag{12}$$

where u denotes a constant drift and e_w^Λ signifies a unit vectors in the wind direction. The addition of S_i, G_i, A_i into Equation (8) modifies the equation of grasshopper motion, which is depicted by the following:

$$X_i = \sum_{\substack{j=1 \\ j \neq i}}^N s(|x_j - x_i|) \frac{x_j - x_i}{d_{ij}} - ge_g^\Lambda + ue_w^\Lambda, \tag{13}$$

where x_i, x_j implies the i th and j th grasshopper and X_i denotes the consecutive location of grasshopper x_i . The grasshoppers accomplish the comfort zone using Equation (13). For identifying the convergence of a certain point, the predefined function is enhanced to achieve a closer optimal solution. Consider that X_i^d is the position of grasshopper i in the d th dimension. Henceforth, the enhanced function is expressed as follows:

$$X_i^d = c_1 \left(\sum_{\substack{j=1 \\ j \neq i}}^N c_2 \frac{ub_d - lb_d}{2} s(|x_j^d - x_i^d|) \frac{x_j - x_i}{d_{ij}} \right) + T_d^\Lambda, \tag{14}$$

where ub_d and lb_d refer to an upper as well as lower bound in the d th dimension, correspondingly. T_d^Λ means the value of the d th dimension. In Equation (14), the gravity factors are fixed to zero and the wind factor often shows a recent best grasshopper. Upon decline, coefficient parameters c_1 and c_2 were employed for simulating the slowdown procedure of grasshoppers that access the food position and utilize the food. While the iterations are enhanced, c_1 is applied for limiting a search scope, whereas c_2 is utilized to reduce the impact of attraction and repulsion among all agents. The maximization function of a variable $c_i (i = 1, 2)$ is provided below.

$$c_i = cMax - l \frac{cMax - cMin}{L}, \tag{15}$$

where $cMax, cMin$ denote the maximum and minimum value of c_1, c_2 , respectively. The parameters are allocated with unique measures, respectively. L shows a high iteration and l is a recent iteration.

3.2. DenseNet Based Feature Extraction Process

The segmented images are fed as input to the DenseNet-201 model. The proficient way to accomplish a prominent outcome in classification issues with a small amount of data is the TL module. Moreover, hyper-tuning of the DTL method is applicable in enhancing the simulation outcome. Here, a DTL approach with DenseNet201 is presented. Therefore, a newly projected approach is applied in feature extraction, where learned weights on the ImageNet dataset and convolutional neural framework are deployed [21]. The framework of the newly developed DTL approach with DenseNet201 for ICH classification is depicted in Figure 2.

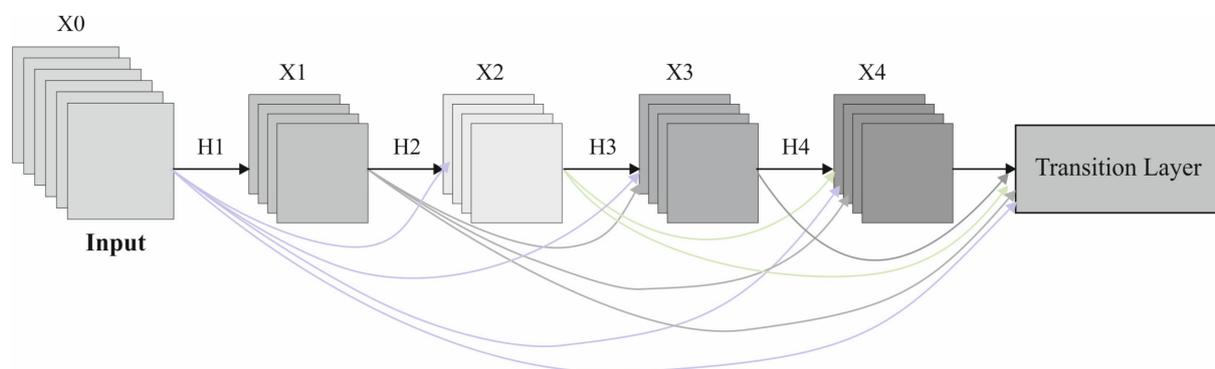


Figure 2. Overall architecture of DenseNet.

DenseNet201 makes use of a condensed network, which provides simple training and efficiency because of the possible feature used for diverse layers, which enhances the difference in the consecutive layer, maximizing the system performance. This method has displayed standard function under different datasets such as ImageNet and CIFAR-100. In order to improve the connectivity in a DenseNet201 scheme, direct communication from previous layers to consecutive layers is employed, as illustrated in Figure 3. The feature combination is expressed in a numerical form:

$$z^l = H_l \left([z^0, z^1, \dots, z^{l-1}] \right) \quad (16)$$

In this approach, H_l means a non-linear transformation described as a composite function with BN, ReLU, and a Conv of (3×3) . $[z^0, z^1, \dots, z^{l-1}]$ represents a feature map combination of equivalent layer 0 to $l - 1$ that has been integrated into a tensor for simple implementation. In the case of the down-sampling mechanism, dense blocks are developed for isolation of layers and transition layers have BN with a 1×1 Conv layer and 2×2 average pooling layer. The progression rate in DenseNet201 defines how a dense structure accomplishes modern intentions for hyper-parameter k . It computes the sufficient progressive rate in which a feature map is assumed as the global state of a system. Thus, a successive layer is composed of feature maps with a previous layer. k feature maps are included to the global state in every layer, in which the overall input feature map at the l th layers $(FM)^l$ is illustrated:

$$(FM)^l = k^0 + k(l - 1) \quad (17)$$

In this framework, the channel in an input layer is referred to as k^0 . In order to enhance the processing efficacy, a 1×1 Conv layer was deployed for all 3×3 Conv layers that mitigates the overall number of input feature maps, which is higher when compared with

output feature maps k . Hence, the 1×1 Conv layer was established, named the bottleneck layer, and it generates $4k$ feature maps.

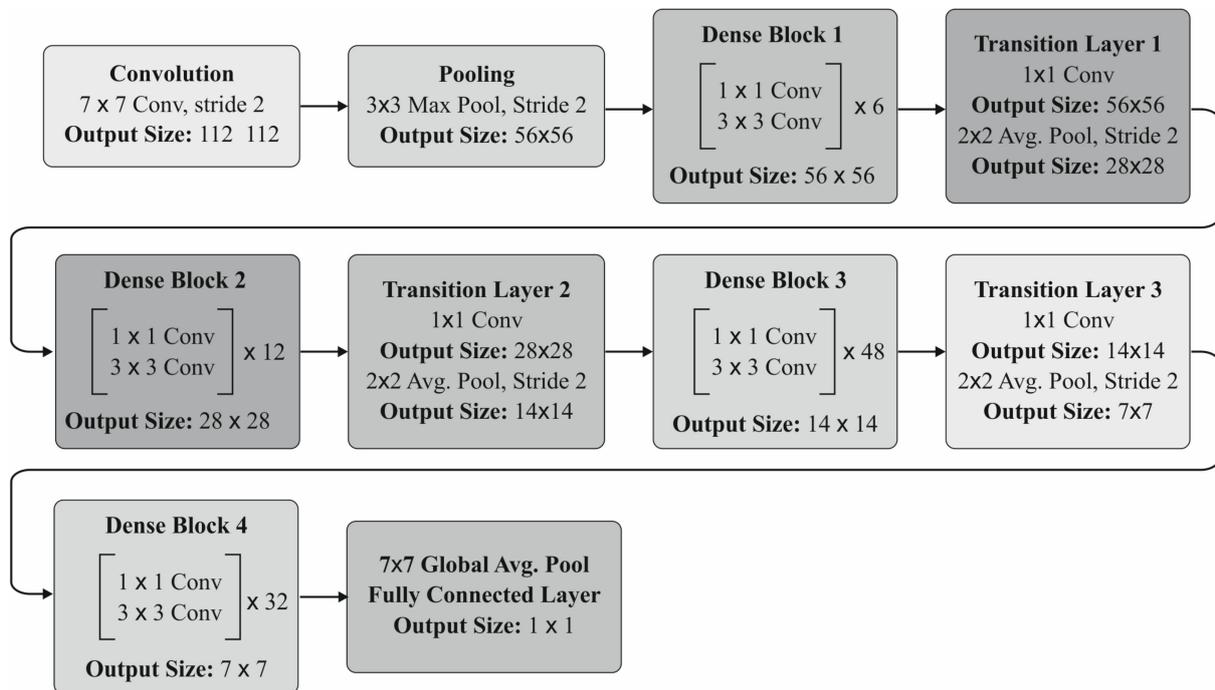


Figure 3. Layered architecture of DenseNet201.

For the purpose of classification [22], two dense layers using neurons were enclosed. The feature extraction system with DenseNet201 and sigmoid activation function is applied for computing binary classifications by inter-changing softmax activation function applied as the traditional DenseNet201 structure. A neuron present in the fully connected (FC) dense layers is linked to all neurons in the former layer. It is defined numerically by FC layer 1, where the input 2D feature map is extended to 1D feature vectors:

$$t^{l-1} = \text{Bernoulli}(p) \tag{18}$$

$$\ddot{x}^{l-1} = t^{l-1} * c^{l-1} \tag{19}$$

$$\ddot{x}^l = f(w^k \ddot{x}^{l-1} + o^l) \tag{20}$$

The Bernoulli function generates a vector t^{l-1} randomly using the 0–1 distribution with a certain probability. c^{l-1} represents the vector dimension. Two layers of the FC layer apply a dropout principle for blocking specific neurons based on the desired probability, which prevents over-fitting problems in a deep system. w^l and o^l describe the weighting as well as offset variables of the FC layer, correspondingly. A sigmoid activation function is applied for changing non-normalized results into binary outputs as zero/one. Henceforth, it is helpful in the consequent classification of ICH positive or negative patients. Here, a sigmoid function is illustrated as follows:

$$y = \frac{1}{1 + e^{-(\sum w_i \cdot x_i)}} \tag{21}$$

where y refers the final outcome of a neuron. w_i and x_i define the weights and inputs, correspondingly.

3.3. ELM-Based Classification Process

After the extraction of a valuable set of feature vectors, the ELM model is applied for the classification method. In general, ELM is defined as a single hidden-layer feed-forward neural network (SLFN). The working principle of SLFN has to be optimized for a system that has to be labelled for data such as threshold value, weight, and activation function; thus, advanced learning is carried out. In the gradient-based learning model, the parameters are modified iteratively to accomplish an optimized measure. Then, with the possibility of a connected device and local minima, the function generates minimal outcomes. In contrast to FNN, it is renewed according to the gradient in ELM; outcomes are estimated, whereas input weights are selected randomly. In the analytic learning process, a success rate is enhanced, as the resolution time and error value mitigate the probability of extracting a local minimum. ELM is also applied for selecting a linear function and enables the cells of the hidden layer, and to apply non-linear (sinusoidal and sigmoid), non-derivatized, or intermittent activation function [23]. Figure 4 showcases the ELM structure.

$$y(p) = \sum_{i=1}^m \beta_i g \left(\sum_{j=1}^n w_{i,j} x_j + b_i \right) \quad (22)$$

where β_i denotes the weights among input and hidden layers and β_j refers to the weight from output and hidden layers; b_j implies a thresholding value of neuron in the hidden layer and g is an activation function. The same number of input layer weights ($w_{i,j}$) and bias (b_j) are allocated arbitrarily. Normally, the activation function ($g(\cdot)$) is allocated for the input layer neuron number (n) and hidden-layer neuron value (m). In this approach, these parameters are referred to as an equilibrium that is unified and organized, and the output layer is depicted in Equation (24).

$$H(w_{i,j}, b_j, x_i) = \begin{bmatrix} g(W_{1,1}X_1 + b_1) & \cdots & g(W_{1,m}X_m + b_m) \\ \vdots & \ddots & \vdots \\ g(W_{n,1}X_n + b_1) & \cdots & g(W_{n,m}X_m + b_m) \end{bmatrix} \quad (23)$$

$$y = H\beta \quad (24)$$

In the training procedure, the training error is minimized to a greater extent. Then, the error function of an output Y_p is attained by the original output \hat{Y}_o value in ELM, $\sum_k^s (\hat{Y}_o - Y_p)$ (with "s": no. of training data) $\| \sum_k^s (\hat{Y}_o - Y_p)^2 \|$, which can be reduced. These functions are used to accomplish output Y_p , achieved by the original value Y_o , which has to be similar to Y_p . While satisfying this function, an unknown parameter in Equation is depicted. The H matrix is defined as a matrix with a lower possibility, which refers to the count of data in the trained set not being equal to the count of features.

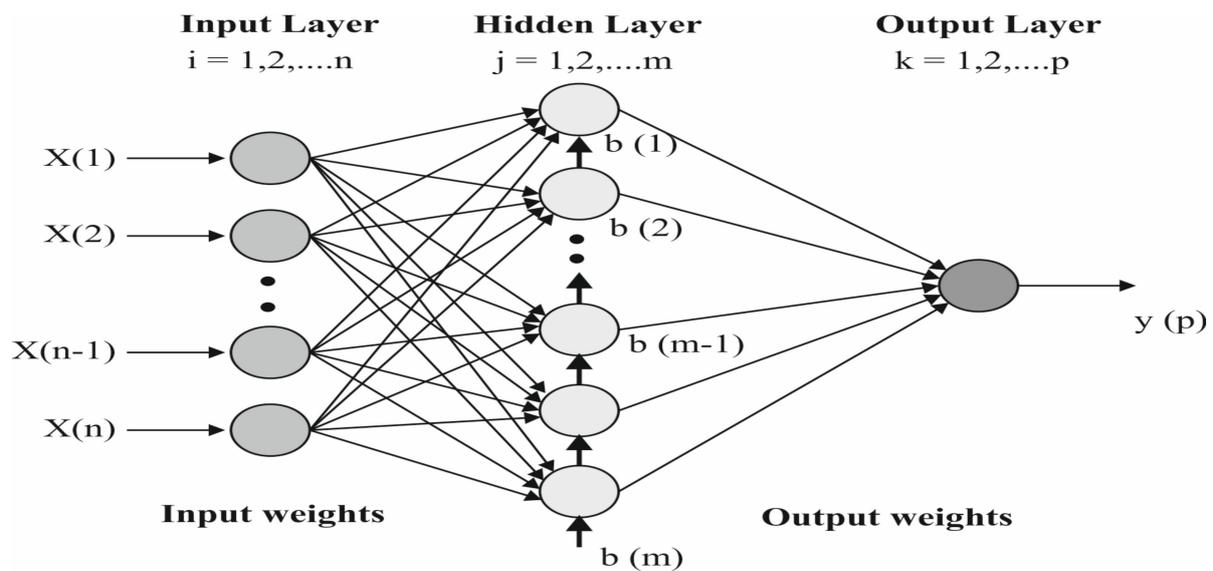


Figure 4. Structure of ELM.

4. Experimental Validation

4.1. Implementation Setup

The proposed DN-ELM model is simulated using the Python 3.4.5 tool. It is executed on a PC motherboard—MSI Z370 A-Pro, processor—i5-8600k, graphics card—GeForce 1050Ti 4GB, RAM—16 GB, OS storage—250 GB, and SSD file storage—1 TB HDD. The parameter settings of the DL-ELM technique are as follows: batch size: 500, max. epochs: 15, dropout rate: 0.2, learning rate: 0.05, weight decay: 0.0001, and momentum: 0.9.

To estimate the performance of the projected method, a research study was carried out utilizing the standard ICH dataset [24]. The dataset includes ICH masks and CT scans, in JPG and NIfTI format, at PhysioNet repository. NIfTI is a type of file format for neuroimaging, which is used very commonly in imaging informatics for neuroscience and even neuroradiology research. The datasets are gathered from the CT scans of 82 individuals under the age group up to 72 years. Furthermore, the dataset has images falling under six classes, namely, intraventricular, with 24 slices; epidural, with 182 slices; intraparenchymal, with 73 slices; subdural, with 56 slices; subarachnoid, with 18 slices; and no hemorrhage, with 2173 slices. For experimental validation, we have used fivefold cross validation for splitting the dataset into testing and training sets.

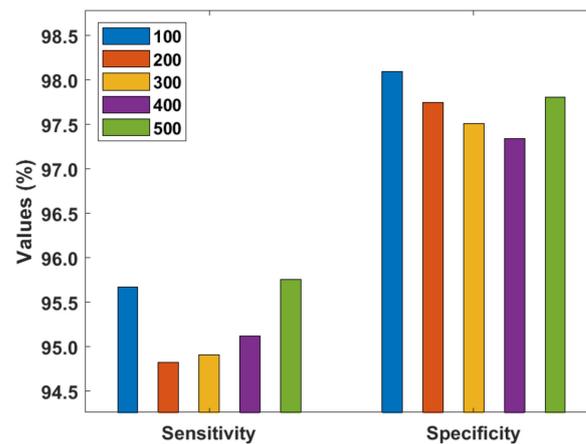
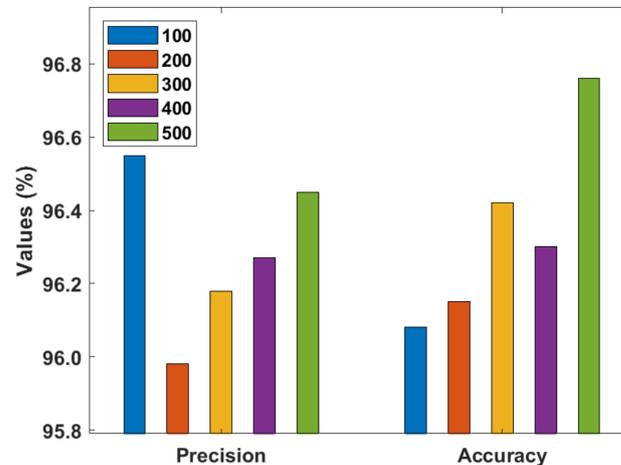
The result is observed in terms of four measures, namely, $sens_y$, $spec_y$, acc_y , and $prec_s$. For comparison functions, the set of methods employed for comparisons are U-Net, Window Estimator Module to a Deep Convolutional Neural Network (WEM-DCNN) [25], Watershed Algorithm with ANN (WA-ANN) [26], ResNext [27], SVM, and CNN approaches.

4.2. Results and Discussion

An analysis of ICH diagnoses results achieved by the DN-ELM model is examined under a different number of epochs. The presented DN-ELM model displayed excellent results under all the distinct epoch counts, as shown in Table 1 and Figures 5 and 6. For example, in the presence of 100 epochs, the DN-ELM model resulted in a $sens_y$ of 95.67%, $spec_y$ of 98.10%, $prec_s$ of 96.55%, and acc_y of 96.08%. Simultaneously, in the presence of 200 epochs, the DN-ELM method resulted in a $sens_y$ of 94.82%, $spec_y$ of 97.75%, $prec_s$ of 95.98%, and acc_y of 96.15%. Concurrently, in the presence of 300 epochs, the DN-ELM approach resulted in a $sens_y$ of 94.91%, $spec_y$ of 97.51%, $prec_s$ of 96.18%, and acc_y of 96.42%.

Table 1. Result analysis of the proposed DN-ELM method on various epoch sizes.

No. of Epochs	$sens_y$	$spec_y$	$prec_s$	acc_y
Epoch-100	95.67	98.10	96.55	96.08
Epoch-200	94.82	97.75	95.98	96.15
Epoch-300	94.91	97.51	96.18	96.42
Epoch-400	95.12	97.34	96.27	96.30
Epoch-500	95.76	97.81	96.45	96.76
Average	95.26	97.70	96.29	96.34

**Figure 5.** Sensitivity and specificity of the DL-ICH model.**Figure 6.** Precision and accuracy analysis of the DL-ICH (deep learning intracranial hemorrhage) model.

In addition, in the presence of 400 epochs, the DN-ELM methodology resulted in a $sens_y$ of 95.12%, $spec_y$ of 97.34%, $prec_s$ of 96.27%, and acc_y of 96.30%. Besides, in presence of 500 epochs, the DN-ELM model resulted in a $sens_y$ of 95.76%, $spec_y$ of 97.81%, $prec_s$ of 96.45%, and acc_y of 96.76%. The average results analysis of the DN-ELM model shown that the DN-ELM technique attained the highest $sens_y$ of 95.26%, $spec_y$ of 97.70%, $prec_s$ of 96.29%, and acc_y of 96.34%.

Figure 7 investigates the results analysis of the DL-ICH approach with the current technique with respect to several performance measures. It is shown that the WA-ANN model provided ineffective ICH diagnosis results by offering a minimum $spec_y$ of 70.13% and $sens_y$ of 60.18%. At the same time, the U-Net algorithm attempted to showcase a slightly higher $sens_y$ of 63.1% and $spec_y$ of 88.6%. Furthermore, the SVM model demonstrated manageable performance with a $sens_y$ of 76.38% and $spec_y$ of 79.41%. In line with this, the WEM-CNN

technique depicted outstanding performance, with a $sens_y$ of 83.33% and $spec_y$ of 97.48%. Moreover, the CNN model provided somewhat higher performance with a $sens_y$ of 87.06% and $spec_y$ of 88.18%. Although the ResNexT model resulted in a competitive $sens_y$ of 88.75% and $spec_y$ of 97.7%, the proposed DN-ELM system achieved a superior ICH diagnostic outcome with a $sens_y$ of 95.26% and $spec_y$ of 97.7%. It is shown that the WA-ANN method provided ineffective ICH diagnosis results by offering a minimum $prec_s$ of 70.08% and acc_y of 69.78%. Simultaneously, the SVM model attempted to demonstrate a somewhat superior $prec_s$ of 77.53% and acc_y of 77.32%. In line with this, the CNN approach portrayed manageable performance with an acc_y of 87.56% and $prec_s$ of 87.98%. At the same time, the U-Net approach displayed even more optimal outcomes with an acc_y of 87% and $prec_s$ of 88.19%. Besides, the WEM-DCNN approach provided slightly higher performance with a $prec_s$ of 89.9% and acc_y of 88.35%. Although the ResNexT technique resulted in a good $prec_s$ of 95.2% and acc_y of 89.3%, the presented DN-ELM method attained optimal ICH diagnostic results, with a $prec_s$ of 96.29% and acc_y of 96.34%.

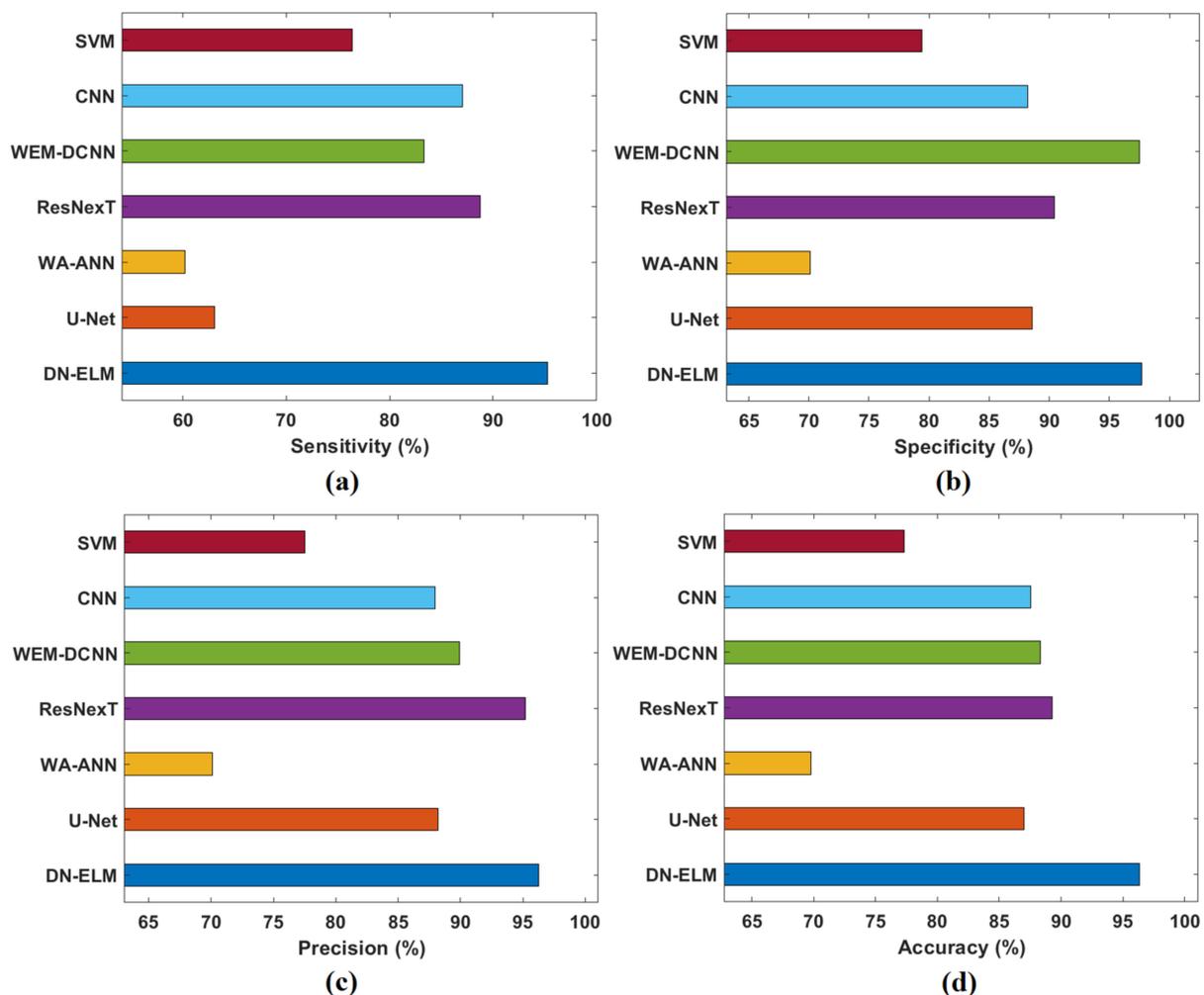
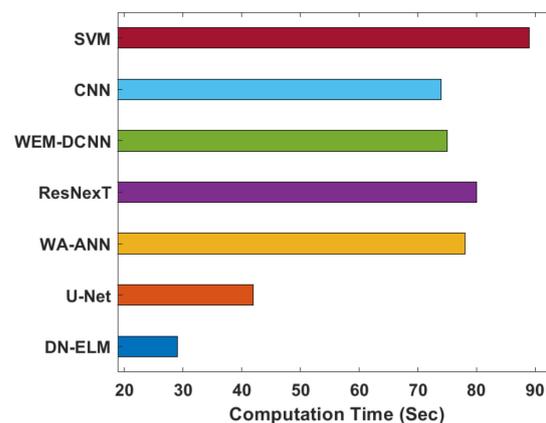


Figure 7. Comparative results analysis of DN-ELM with existing models: (a) sensitivity, (b) specificity, (c) precision, and (d) accuracy.

Table 2 and Figure 8 present the analysis of the results offered by the DN-ELM with the existing models in terms of computation time (CT). The experimental outcome specified that the SVM technique demonstrated inferior results, with a higher CT of 89 s. Furthermore, the ResNexT and WA-ANN models demonstrated lower CTs of 80 s and 78 s, respectively.

Table 2. Result analysis of the proposed DN-ELM model with existing methods in terms of computation time.

Methods	Computation Time (Sec)
DN-ELM	29.00
U-Net	42.00
WA-ANN	78.00
ResNexT	80.00
WEM-DCNN	75.00
CNN	74.00
SVM	89.00

**Figure 8.** Computation time analysis of the DN-ELM model.

In line with this, the CNN and WEM-DCNN methods demonstrated moderate CTs of 75 s and 74 s correspondingly. Besides, the U-Net model displayed even better performance, with a CT of 42 s, whereas the DN-ELM technique attained superior results, with a minimum CT of 29 s. The experimental outcome ensured the outstanding performance of the DN-ELM system with the existing methods.

5. Conclusions

This paper introduced a new DL-ELM technique for the diagnosis and classification of ICH. The presented method comprises several subprocesses, such as classification, preprocessing, segmentation, and feature extraction. The DL-ELM model undergoes a preprocessing step, where the input data from the NIfTI file are transformed into JPEG format. Then, the TEGOA technique is employed for the image segmentation process. The application of GOA helps to determine the optimal threshold value to perform multilevel thresholding-based image segmentation. Furthermore, the segmented image is fed as input to the DenseNet-201 model. Subsequent to the extraction of a valuable set of feature vectors, the ELM model is employed for the classification process. A detailed experimental results analysis takes place to determine the performance of the DL-ELM approach. The outcome of the simulations implied that the DN-ELM model outperformed all of the state-of-the-art ICH approaches, with a $sens_y$ of 95.26%, $spec_y$ of 97.70%, $prec_s$ of 96.29%, and acc_y of 96.34%. As a part of the future scope, the hyper parameters of the DenseNet methodology should be determined using the bio-inspired optimization algorithms to improve the classification outcome.

Author Contributions: Conceptualization, S.S.; Formal analysis, J.J.A.; Methodology, S.G.; Supervision, V.V.; Writing—review & editing, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Qureshi, A.I.; Mendelow, A.D.; Hanley, D.F. Intracerebral haemorrhage. *Lancet* **2009**, *373*, 1632–1644. [[CrossRef](#)]
2. Mayer, S.A.; Kreiter, K.T.; Copeland, D.; Bernardini, G.L.; Bates, J.E.; Peery, S.; Claassen, J.; Du, Y.E.; Connolly, E.S. Global and domain-specific cognitive impairment and outcome after subarachnoid hemorrhage. *Neurology* **2002**, *59*, 1750–1758. [[CrossRef](#)] [[PubMed](#)]
3. Hylek, E.M.; Singer, D.E. Risk factors for intracranial hemorrhage in outpatients taking warfarin. *Ann. Intern. Med.* **1994**, *120*, 897–902. [[CrossRef](#)] [[PubMed](#)]
4. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
5. Cao, G.; Wang, Y.; Zhu, X.; Li, M.; Wang, X.; Chen, Y. Segmentation of intracerebral hemorrhage based on improved U-Net. In Proceedings of the 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), Shenyang, China, 11–13 December 2020; pp. 183–185.
6. Dong, J.; Shi, F. Multi-dimensional data analysis of intracerebral hemorrhage from CT images. In Proceedings of the 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China, 16–18 October 2010; pp. 406–409.
7. Davis, V.; Devane, S. Diagnosis & classification of brain hemorrhage. In Proceedings of the 2017 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 1–2 December 2017; pp. 1–6.
8. Shanbhag, S.S.; Udipi, G.R.; Patil, K.M.; Ranganath, K. Analysis of brain MRI images of intracerebral haemorrhage using frequency domain technique. In Proceedings of the 2011 International Conference on Image Information Processing, Shimla, India, 3–5 November 2011; pp. 1–5.
9. Amir, N.S.B.S.; Chellappan, K.; Kang, L.Z.; Mukari, S.; Sahathevan, R. MR image enhancement for ICH classification. In Proceedings of the 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia, 4–7 December 2016; pp. 160–165.
10. Ghafaryasl, B.; van der Lijn, F.; Poels, M.; Vrooman, H.; Ikram, M.A.; Niessen, W.J.; van der Lugt, A.; Vernooij, M.; de Bruijne, M. A computer aided detection system for cerebral microbleeds in brain MRI. In Proceedings of the 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2–5 May 2012; pp. 138–141.
11. Yuh, E.L.; Gean, A.D.; Manley, G.T.; Callen, A.L.; Wintermark, M. Computer-Aided Assessment of Head Computed Tomography (CT) Studies in Patients with Suspected Traumatic Brain Injury. *J. Neurotrauma* **2008**, *25*, 1163–1172. [[CrossRef](#)] [[PubMed](#)]
12. Li, Y.; Wu, J.; Li, H.; Li, D.; Du, X.; Chen, Z.; Jia, F.; Hu, Q. Automatic Detection of the Existence of Subarachnoid Hemorrhage from Clinical CT Images. *J. Med. Syst.* **2012**, *36*, 1259–1270. [[CrossRef](#)] [[PubMed](#)]
13. Kuo, W.; Häne, C.; Yuh, E.; Mukherjee, P.; Malik, J. Cost-Sensitive active learning for intracranial hemorrhage detection. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*; Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 715–723.
14. Prevedello, L.M.; Erdal, B.S.; Ryu, J.L.; Little, K.J.; Demirel, M.; Qian, S.; White, R.D. Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. *Radiology* **2017**, *285*, 923–931. [[CrossRef](#)] [[PubMed](#)]
15. Chilamkurthy, S.; Ghosh, R.; Tanamala, S.; Biviji, M.; Campeau, N.G.; Venugopal, V.K.; Mahajan, V.; Rao, P.; Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* **2018**, *392*, 2388–2396. [[CrossRef](#)]
16. Ye, H.; Gao, F.; Yin, Y.; Guo, D.; Zhao, P.; Lu, Y.; Wang, X.; Bai, J.; Cao, K.; Song, Q.; et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. *Eur. Radiol.* **2019**, *29*, 6191–6201. [[CrossRef](#)] [[PubMed](#)]
17. Grewal, M.; Srivastava, M.M.; Kumar, P.; Varadarajan, S. RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 281–284.
18. Lee, H.; Yune, S.; Mansouri, M.; Kim, M.; Tajmir, S.; Guerrier, C.E.; Ebert, S.A.; Pomerantz, S.R.; Romero, J.M.; Kamalian, S.; et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat. Biomed. Eng.* **2018**, *3*, 173–182. [[CrossRef](#)] [[PubMed](#)]
19. Jnawali, K.; Arbabshirani, M.R.; Rao, N.; Patel, A.A. Deep 3D convolution neural network for CT brain hemorrhage classification. In *Medical Imaging 2018: Computer-Aided Diagnosis*; SPIE: Washington, DC, USA, 2018; Volume 10575, p. 105751C.
20. Rajinikanth, V.; Satapathy, S.C.; Fernandes, S.L.; Nachiappan, S. Entropy based segmentation of tumor from brain MR images—a study with teaching learning based optimization. *Pattern Recognit. Lett.* **2017**, *94*, 87–95. [[CrossRef](#)]
21. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
22. Jaiswal, A.; Gianchandani, N.; Singh, D.; Kumar, V.; Kaur, M. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5682–5689. [[CrossRef](#)] [[PubMed](#)]
23. Toprak, A. Extreme Learning Machine (ELM)-Based Classification of Benign and Malignant Cells in Breast Cancer. *Med. Sci. Monit.* **2018**, *24*, 6537–6543. [[CrossRef](#)]
24. Hssayeni, M.D.; Croock, M.S.; Salman, A.D.; Al-Khafaji, H.F.; Yahya, Z.A.; Ghoraani, B. Intracranial Hemorrhage Segmentation Using a Deep Convolutional Model. *Data* **2020**, *5*, 14. [[CrossRef](#)]

25. Elena, S.; Gina, S.; Sarina, F.; Danielle, M.; Twinkle, J.; Joseph, K.T.; Jickling, G.C. Hemorrhagic Transformation in Ischemic Stroke and the Role of Inflammation. *Front. Neurol.* **2021**, *12*, 597.
26. Danilov, G.; Kotik, K.; Negreeva, A.; Tsukanova, T.; Shifrin, M.; Zakharova, N.; Batalov, A.; Pronin, I.; Potapov, A. Classification of Intracranial Hemorrhage Subtypes Using Deep Learning on CT Scans. *Stud. Health Technol. Inform.* **2020**, *272*, 370–373. [[PubMed](#)]
27. Karki, M.; Cho, J.; Lee, E.; Hahm, M.-H.; Yoon, S.-Y.; Kim, M.; Ahn, J.-Y.; Son, J.; Park, S.-H.; Kim, K.-H.; et al. CT window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings. *Artif. Intell. Med.* **2020**, *106*, 101850. [[CrossRef](#)]