Article

Construction of VGG16 Convolution Neural Network (VGG16_CNN) Classifier with NestNet-Based Segmentation Paradigm for Brain Metastasis Classification

Abdulaziz Alshammari

Abstract: Brain metastases (BMs) happen often in patients with metastatic cancer (MC), requiring initial and precise diagnosis of BMs, which remains important for medical care preparation and radiotherapy prognostication. Nevertheless, the susceptibility of automated BM (ABMS) diagnosis is unfairly great for minute BMs, and integrating into medical exercises to distinguish true metastases (MtS) from false positives remains difficult. For enhancing BM classification execution, MtS localization is performed through the NestNet framework. Subsequent to segmentation, classification is performed by employing the VGG16 convolution neural network. A novel loss function is computed by employing the weighted softmax function (WSF) for enhancing minute MtS diagnosis and for calibrating susceptibility and particularity. The aim of this study was to merge temporal prior data for ABMS detection. The proffered VGG16_CNN is capable of differentiating positive MtS among MtS candidates with high confidence, which typically needs distinct specialist analysis or additional investigation, remaining specifically apt for specialist reinforcement in actual medical practice. The proffered VGG16_CNN framework can be correlated with three advanced methodologies (moU-Net, DSNet, and U-Net) concerning diverse criteria. It was observed that the proffered VGG16_CNN attained 93.74% accuracy, 92% precision, 92.1% recall, and 67.08% F1-score.

Keywords: brain metastases; classification; segmentation; deep learning; neural network; VGG16_CNN; NestNet

1. Introduction

Patients with MC possess an elevated threat of acquiring brain metastases (BMs), having a rough occurrence rate of up to 40%. The advancement of BM causes decreased or nil efficiency of general systemic medical care [1]. Hence, victorious medical care of BMs remains extremely important for a patient’s existence and standard of living. Since whole-brain radiation therapy (WBR) of BMs leads to cognitive deteriorations, stereotactic radiosurgery (SRSy) is attracting increased interest for BM therapy. SRSy dispenses highly focused radiation toward the metastasis areas with less dosage toward the neighboring typical brain tissues. Therefore, this leads to not many side-effects when compared with WBR [2]. For SRSy therapy preparing and scheduling, the BM count, dimension, edge, and position remain significant data that need precise identification for ensuing BM segmentation. Presently, BMs can be detected physically by neuroradiologists and radiation oncologists through a time-consuming process that suffers from inter-rater variability [3]. In particular, some metastases (MtS) are simply disregarded in physical identification since they are positioned only in some image slices and normally possess low contrast. Furthermore, a few anatomical forms such as blood vessels resemble BMs in two-dimensional intersection planes, hindering their detection [4]. Hence, computer-aided automated brain metastasis (ABMS) detection possesses significant medical value. For ABMS identification and segmentation, standard machine learning methodologies such as template matching,
support vector machine, and AdaBoost have been implemented [5]. Nevertheless, these proved mediocre compared to the recent deep learning (DL) methodologies.

Owing to the latest escalation of DL approaches, even though most analysts concentrated on the segmentation of chief brain tumors (BTs) such as gliomas, the study of DL for BM identification and segmentation is improving. Concerning neural network (NN) frameworks for BM segmentation, three-dimensional U-Net and DeepMedic1 remain the preferred networks [6]. Other NNs include GoogLeNet, V-Net, faster R-CNN, single-shot detectors, and custom convolutional NNs (CNNs) [7]. For saving memory, the GoogLeNet technique employs seven slices: a center slice with six nearby slices, resembling a 2.5-dimensional paradigm. The remaining NNs employ three-dimensional sub-volumes for training [8]. To train segmenting NNs, loss functions (LFs) such as binary cross-entropy (BCE), Dice similarity coefficient (DSC), and intersection over union (IOU) are normally implemented. For enhancing BM segmentation, novel LFs can be proffered. Herein, the contribution of this process is ensuring that the network employs its whole receptive field. While conducting training, the extra outputs are employed as supplementary loss layers by correlating them to inferior variants of the reference label data. The aim of this study was to differentiate true positive (TP) MtS among MtS candidates (MCs) with high confidence, which typically requires unique specialist analysis or additional investigation, remaining specifically apt for specialist reinforcement in actual medical exercise. VGG16_CNN attains great susceptibility for BM identification and differentiates MtS among MCs with high confidence.

The remainder of this paper is structured as follows: Section 2 highlights some existing studies, Section 3 presents the background of prediction criteria for BM classification, Section 4 describes the proffered technique and methods, Section 5 presents the experimental results and discussion, and Section 6 provides a conclusion and insights for future study.

2. Associated Studies

Tumors and neoplasms are created when cells in any part of the body develop abnormally and form a mass. There are two different kinds of tumors: one is benign and the other is cancerous or "malignant". A benign tumor does not represent a high danger to the body because it does not spread to other parts of the body. A malignant tumor, however, is extremely dangerous as it spreads to other body organs [9]. This section addresses the associated literature and conveys the limitations addressed by the proffered methodology. Generally, similar to this study, many studies concentrated upon diverse kinds of MtS classification.

The authors of [10] introduced a new three-dimensional multi-attention guided multi-task learning network for concurrent gastric tumor segmentation with lung nodule classification that achieves complete utilization of the corresponding data with disparate sizes, scales, and tasks. In particular, the task correspondence and heterogeneity of the CNN comprise scale-aware attention-guided shared feature learning (AGSFL) for refinement and global multiscale (MS) features, and task-aware AGSFL for task-particular discriminatory features. The major applications of convolutional neural networks are image recognition and classification. It is also the only use case involving the most advanced frameworks in medical imaging.

The authors of [11] proposed a few-shot learning methodology for classifying an image patch that comprises tumor cells. Specifically, a patch-level unadministered cell scoring technique is presented that solely depends upon images having restricted labels. The chief notion of this methodology remains that, while clipping patch A from whole-slide images (WSI), in addition to a sub-patch B from patch A, A’s cell count remains constantly above that of B. Accordingly, the unlabeled images can be employed for learning the cell counting data for extracting notable features.

The authors of [12] presented a new deep regional MtS segmentation (DRMS) architecture for lymph node status classification. Initially, a deep segmentation network (DSN) was
proffered for identifying the regional MtS within the patch level. Next, the density-based spatial clustering of applications with noise (DBSCAN) was embraced for prognosticating the complete MtS from independent slides. Lastly, pN phases in patients were decided by totaling independent slide-level prognoses.

The moU-Net performed well for the segmentation of glioblastoma and low-rank glioma [13]. With similar network encoder and decoder layers to the conventional U-Net, moU-Net includes an extra output layer toward the decoder. U-Net is marked by an encoder that extracts lower-grade portrayals of the input data (ID) and is associated with a decoder that rebuilds a correlated label map with skip connections between intermediate phases of the two units. Since its initial proposal, many augmentations concerning the network framework and the training procedure have been applied to U-Net [14].

The authors of [15] computed the susceptibility and particularity of the paradigm for each pel, as opposed to lesion, greatly benefiting bigger lesions. The utilized database also possesses a common dispensation of primary cancer (PC) kinds that are employed for exhibiting the DCNN’s strong execution for disparate chief tumors. The authors were certain that their work remained a rational and, hence, medically pertinent observance of DCNN’s anticipated execution.

The ideal therapy of patients with BMs relies upon the intended condition. Frequent operation, radiation operation, whole-brain (WH) radiation therapy, and chemotherapy can be employed together for attaining lengthier remissions and ideal symptom relief. Patients with several MtS generally obtain WB irradiation (WBI). It remains certainly arduous to seek the MtS degree using the aforementioned methodologies. Thus, a new classifier is proffered in this study.

In spite of encouraging classification findings in many evaluated deep learning studies, care is urged because overfitting must be avoided when using big, annotated, and good datasets [16]. Three studies [17–19] tested ensemble learning techniques. In order to create a model that outperforms individual classifiers, ensemble learning combines many algorithms that are trained on either various datasets or the same data, but with distinct algorithms.

3. Background of Predictive Attributes for BM Classification

The prognosis of patients with BMs is normally poor and notably ruins quality of life. The median survival of patients with BMs does not exceed 5 months in spite of the present therapeutic possibilities. However, the description of perfectly identified predictive subsets remain vital for the selection of a customized treatment plan. It remains significant to detect subsets of patients with suitable predictive attributes, which can then lead to the development of therapy targeting survival and an improved quality of life [20]. Conversely, for patients devoid of ideal extracranial illness control and/or comorbidities (which can restrict the tolerance of fierce therapies), the therapy goal remains to stabilize BMs for administering symptoms and restricting malignancy.

Significant predictive attributes include key performance status (KPS), the BM count, the lack of systemic MtS, principal tumor control, and age. The Radiotherapy Oncology Group (RTOG) modeled a prediction ranking scheme inferred from the assessment of individual predictive attributes for patients with BMs and proposed three subsets (RPA classes) [21]. This led to the establishment of a novel and elaborate predictive index (Graded Prognostic Assessment, GPA), which considers four factors: age, KPS, the existence of extracerebral MtS, and BM count [22].

4. System Paradigm

A comprehensive schematic of the proposed BM classification employing the BT database is illustrated in Figure 1. The BT dataset comprised 3064 T1-weighted contrast-optimized images. The database was then preprocessed, rescaled, and generalized. Next, segmentation was performed and MtS identification was carried out by employing the
NestNet (NtNt) framework. Subsequent to segmentation, classification was performed by employing the VGG16 convolution NN.

**Figure 1.** Comprehensive framework for BM classification.

### 4.1. Image Initialization

The image data were retroactively recovered from our institutionalized e-clinical registers and anonymized prior to the assessment. Axial three-dimensional spoiled gradient recall or magnetization-prepared rapid gradient echo T1-weighted contrast-augmented MR images were recovered for assessment. Patients with single or multiple BMs were chosen consecutively while conducting the medical readout. MRI was performed using 1.5 T and 3 T scanners from two chief merchants. Overall, data were recovered for 121 patients with 2053 MtS, resulting in 14,350 image slices from a total of 361 scans. Patients were incorporated into the research for definitive medical and imaging detection of BMs. Patients with ambiguous lesions were not considered. Pathological verification of BMs through brain biopsy or resection was performed in 43 patients (36%). In the remaining 78 patients (64%), brain operation was not performed; thus, pathologic detection using brain tissue was unavailable. However, as the doctors did not approve the risk of brain surgery in consideration of the BMs identified through medical and imaging data, all 78 patients possessed a noted detection of PC elsewhere in the body. Accordingly, the team unanimously agreed to the medical detection of BMs in these patients. The majority of patients completed several scans, whereas some completed a single scan.

### 4.2. Brain Image Preprocessing

Prior to employing images for training or testing, a sequence of procedures was applied for image preprocessing. Firstly, since the initial images had disparate matrix dimensions, the images were padded with zeroes. Scaling was carried out next, if required, for transforming entire images to achieve similar matrix dimensions of $900 \times 900$. For the reference image (RI), the histogram consisted of homogeneous low-intensity regions of interest (ROIs) (low-intensity region—LIR) and high-intensity ROIs (high-intensity region—HIR). The histogram ranged from LIR to HIR illumination levels. The image intensities were mapped as values between HIR and LIR; the former was considered the value at the maximal decile, while the latter was considered the value at the minimal decile. This
assisted in eliminating background noise and outliers as defined above according to the following equation:

$$f'(x, y, z) = \frac{f(x, y, z) - LIR}{HIR - LIR}$$

where $f(x, y, z)$ indicates the initial reference gray value at $(x, y, z)$, and $f'(x, y, z)$ indicates the corresponding transitioned grayscale value. Subsequent to primary scaling, histogram normalization can be performed. The ROI histogram is elongated and transformed to cover all grayscale levels (GL) within the input image (IpI) as follows:

$$g'(x, y, z) = \frac{HIR - LIR}{S_{\text{max}} - S_{\text{min}}} (g(x, y, z) - S_{\text{min}} + LIR).$$

When the IpI’s target histogram $g(x, y, z)$ begins at $S_{\text{min}}$ and elongates to $S_{\text{max}}$ GL within the ROI, the image can be scaled between the bottom edge $m_1$ and the top edge $m_2$, whereby the voxels within the novel normalized image $g'(x, y, z)$ are between a minimal degree (LIR) and a maximal degree (HIR). The resulting variables $m_1$ and $m_2$ remain the bottom edge and top edge of the ROI before scaling up. The normalization action is expressed as $N(x, y, z)$ according to the following equation:

$$N(x, y, z) = \begin{cases} 
\mu_s + (g(x, y, z) - \mu_i) \frac{\text{LIR} - \mu_i}{S_{\text{max}} - \mu_i}, & m_1' \leq g(x, y, z) \leq \mu_i \\
\mu_s + (g(x, y, z) - \mu_i) \frac{\text{HIR} - \mu_i}{S_{\text{max}} - \mu_i}, & \mu_i \leq g(x, y, z) \leq m_2'
\end{cases}$$

where $\mu_i$ and $\mu_s$ represent the average values for the IpI and ROI histograms, respectively. $S_{1i}$ and $S_{2i}$ represent the voxel values of the IpI.

### 4.3. MtS Localization Employing NtNt Framework

The preprocessed images incorporated the following localization-related features:

- Left/right hemisphere or central form,
- Cerebral lobes: frontal, parietal, temporal, and occipital lobe,
- Insular cortex,
- Subcortical forms: basal ganglia, thalamus, brainstem, corpus callosum, and cerebellum,
- Eloquent brain regions: vision center (the region surrounding Sulcus calcarinus), auditory center (Gyrus temporales transversae), Wernicke’s region (from Gyrus temporalis superior’s dorsal area to the parietal lobe’s Gyrus angularis et supramarginal), Broca’s region (Gyrus frontalis inferior’s Pars triangularis et opercularis), the primary somatosensory cortex (Gyrus postcentralis), and primary somatomotor cortex (Gyrus praecentralis).

The localization-related features were not considered as cooperatively unique, whereby each lesion was ascribed to several of the abovementioned classes, e.g., left hemisphere, temporal lobe, and vision center.

Pairwise differences (PDs) were calculated for each lesion’s binary feature vectors (FVs), as well as histological subtypes, by exploring the Jaccard distance $= (1 - \text{Jaccard index})$. The consequential distance matrix (DM) (size $239 \times 239$) was clustered by employing agglomerative stratified clustering with mean linkage. Additionally, PDs were encapsulated over entire lesions (by taking the mean) within a histological tumor type to yield the DM for each type. The resulting DM was transitioned into an undirected graph’s (UG) affinity matrix by calculating its entries $a_{ij} = 1 - di, j$ with $di, j$ remaining the DM’s $(i, j)$-th entry. Lastly, the UG could be viewed by employing a spectral layout with high-affinity nodes closer than low-affinity nodes.

### 4.4. MtS Area Segmentation Employing NtNt Framework

To effectively perform image segmentation, this study proposes a novel DL method consisting of an encoding module (EU) in which two disparate timespans can be employed as the input for executing feature extraction (FE). This unit contains four stratified ranks,
and feature tensors in a similar rank contain a similar length and breadth. Nevertheless, the channels remain disparate. The method also consists of a decoding module which segments the extracted features and upsamples them toward an image with similar length and breadth to the change-labeled image. The NtNt framework consists of two corresponding units for processing images of disparate timespan.

The segmentation procedure framework is illustrated in Figure 2, where Figure 2a depicts the downsampling module (DM) [23]. Various timespan images can be employed as the ID for this DM. Subsequent vectors within the FV of a similar level can be attained from the previous vector (PV) following three procedures: ReLU, Conv (red arrows), and splicing of entire PVs and next-level vectors (NLVs) after upsampling (green arrows). NLVs can be downsampled from the previous-level vector (PLV; blue arrow). Figure 2b depicts the upsampling module. A dense framework can be employed for the top three layers, and various level vectors can be linked via the upsampling procedure. Here, four outcomes from O1 to O4 can be ultimately attained, which are later joined using a concatenating procedure and transitioned to O5 via a convolution procedure.

Figure 2. Segmentation procedure employing NtNt framework.

4.5. Encoding Module

The feature extraction units target multiscale convolutional features in images X and Y. \( F_{x/y}^{i,j} \) is employed to differentiate the features from X and Y. \( F_{x/y}^{i,j} \) is computed using Equation (4).

\[
F_{x/y}^{i,j} = c \left( a \left( F_{x/y}^{i,j}, u F_{x/y}^{i+1,j+1} \right) \right).
\]  

(4)

Considering \( j \in \{ (0,1), (0,2), (1,1), (2,1) \} \) and \( F_{x/y}^{i,j} = d \left( F_{x/y}^{i-1,j} \right), j \in \{ 1, 0, (2,0), (2,0), (4,0) \} \). In Equation (4), \( F_{x/y}^{i,j} \) and \( F_{x/y}^{i,j} \) portray the features from images X and Y, \( c(\cdot) \) indicates the concat procedure, which concatenates the features by channels, \( u(\cdot) \) indicates
the upsample procedure, which upsamples the feature's length and breadth to half of the initial dimension, \( a(\cdot) \) indicates the EM’s calculation, and \( x/y \) indicates \( X \) or \( Y \). ADO is employed for synthesizing the features from images \( X \) and \( Y \). \( F^{i,j}_{x/y} \) as the ADO’s outcome, can be computed using Equation (5).

\[
F^{i}_{D} = \beta\left(F^{a,b}_{x}, F^{a,b}_{y}\right).
\]  

Considering \( i \in 1, 2, 3 \ldots 9 \), \( F^{a,b}_{x} \) and \( F^{a,b}_{y} \) portray the features from images \( X \) and \( Y \). \( \beta(\cdot) \) denotes the ADO described in Equation (1). The EM remains a dual framework, in which two alike units are employed for feature extraction, feature downsampling, eigenvalue activation, and stitching of the disparate levels of feature layer on the bitemporal images. The feature data from two images can be extracted by EM.

### 4.6. Decoding Module

The outcomes of the initial elements \( \left(D^{1} - D^{9}\right) \) are considered as the downsampling module input. In the DM, similar level features are linked to the remaining features, and low-level features are upsampled and concatenated to the previous level’s features. \( F^{i,j}_{x/y} \) is employed to portray the value of \( XY^{i,j} \), where \( F^{i,j}_{x/y} \) is computed as follows:

\[
F^{i,j}_{x/y} = c\left(u\left(F^{a,b}_{x,y}\right), F^{i}_{D}\right).
\]  

A dense block unit is employed in the downsampling module’s top three layers for feature fusion to enhance the effectiveness of feature usage. For each layer, the feature maps of the previous layer are employed as the input. In line with the remaining transition identification network units, the NiNt framework’s chief disparities incorporate the sima network to execute segmentation of two individual images. To resolve the issue where the transformed and untransformed pels remain unbalanced, the LF employed in this study comprises balanced binary cross-entropy loss (BBCEL) and Dice loss. The entire layer result at level 0 computes the loss with a disparate weight. The comprehensive LF is described as

\[
L = \sum_{i=1}^{x} \omega_{i}L_{side}^{i},
\]  

where \( \omega_{i} \) portrays the i-th output layer’s weight. \( L_{side}^{i} \) comprises BBCEL and DL, which can be described by Equation (8).

\[
L_{side}^{i} = L_{BCE}^{i} + L_{dice}^{i}.
\]  

Regardless of the training database or confirmation database, most images only possess a tiny area that remains transformed. A few LFs employed in segmentation do not perform well in transition identification. Thus, the balanced BBCEL is considered in the LF portion, which can be described by Equation (9).

\[
L_{BCE} = \sum_{n=1}^{N} w_{n}[y_{n} \times \log x_{n} + (1 - y_{n}) \times \log (1 - x_{n})],
\]  

where \( y_{n} \) indicates the ground truth of pel \( n \) (zero or one), and \( x_{n} \) indicates the estimated value of pel \( n \). \( w_{n} \) portrays the ratio of the present pel (zero or one) to the sum of pels. \( N \) is the sum of pels.

### 4.7. Classification Employing VGG16 Conv NN

In Dice loss, the CNN or ConvNet denotes the deep NN commonly implemented for assessing an image’s visual qualities. VGG-16 is a CNN mostly employed to process images. VGG-16’s 16 layers consist of convolutional, pooling, and fully connected layers (FCLs).
Figure 3 illustrates the VGG-16 CNN framework. The network’s framework contains multiple convolution layers (CLs) and three FCLs. These FCLs are portrayed as dense layers [24]. Training images are used as input. Patches of these images are extracted and forwarded to the network during training.

![VGG-16 CNN framework](image)

Figure 3. VGG-16 CNN framework.

4.8. Shared Core Layer

The multitask network of this study was constructed using the VGG16 network proposed by Simonyan and Losch [25], constituting a stack of 3 × 3 CLs (Conv) with nonlinearity (ReLU) and 3 × 3 max-pooling layers (MPL) (pool) for downsampling. The framework is reiterated until the result contains a tiny spatial dimension, and a decision is made with respect to the result by the FCLs and the softmax layer to compute class probabilities. As this network consists of five MPLs, the spatial dimension is lessened by a factor of 25 or 32 prior to making it to the FCLs.

4.9. Atrous Convolution Block

The atrous convolution (AC) block classifies the ID into four divisions for corresponding computation. The four divisions employ a similar quantity of convolution kernels (CKs). The initial division employs a conventional 3 × 3 convolution. The next division employs an AC with a CK dimension of 3 × 3 and two dilation rates (DRs). The next division employs an AC with a CK dimension of 3 × 3 and three dilation rates. The last division employs an AC with a CK dimension of 3 × 3 and five DRs. Following the four CLs for classification into divisions, there is a transformed linear module (Relu) and a batch normalization (BN) procedure. Lastly, the concat procedure is employed to concatenate the feature maps of the four divisions. AC can efficiently enlarge the receptive region without affecting computation. The AC block employs convolutions with various dilation rates to concurrently extract images at several scales, facilitating network PE.

4.10. Optimized Boosting Strategy

The boosting strategy optimizes the precision at the start of augmenting the number of classifiers. If a great number of weak classifiers (WC) is employed, this would not continually enhance the precision and would result in overfitting [26]. In this study, exit was observed that three WCs were adequate for creating a robust classifier that considerably enhances the classification’s execution. One-third of the data samples were randomly chosen for training the initial WC VGG16-T. Considering misclassified samples from the previous round as the training data (TD), the increasing weight of erroneous samples could be improved according to criteria for training the subsequent WC VGG16-T to learn many portrayal features, thereby alleviate the unbalance. Many characteristic and discriminatory samples can be chosen for training and testing to augment the method’s strength.
As shown in Figure 4, the chief WC was trained by the filtered databases [27]. The initial error data (misclassified (0)) within the red rectangle box and the novel training data were fused for training the next WC. The below criteria were enhanced by employing erroneous data from the initial two data and the remaining novel data. To decrease the number of false positives (FPs), boosting is an analytical learning methodology commonly employed with optimized classifiers. The classifiers remain linearly joined to enhance classification by assigning disparate weights to training samples (TS).

Figure 4. Weighted boosting methodology’s framework.

4.11. Weighted Softmax Function

The softmax classifier (SC) is a normalization method that splits target variables into multiple classes [28,29]. SC was employed for BM pathological feature identification in CT images.

It can be presumed that there are \( N \) IPIs \( \{x_i, y_i\}_{i=1}^{N} \) in every image according to a tag out of \( \{y_i \in \{1, 2, 3, \ldots, k\}, k > 2\} \). The sum of classes is indicated by \( k \) (\( k = 4 \) in this methodology). For a provided test image \( x_i \), the hypothesis function is to predict the probability value \( p(y_i = j/x_i) \) of every class \( j \). The value of \( h_\theta(x_i) \) can be computed as follows:

\[
    h_\theta(x_i) = \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T x_i)} \begin{bmatrix} p(y_i = 1) & \ldots & x_i; \theta \\ \vdots & \ddots & \vdots \\ p(y_i = k) & \ldots & x_i; \theta \end{bmatrix}.
\]

where \( \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T x_i)} \) portrays the probability dispensation’s normalization, i.e., the sum of entire probabilities remains 1, and \( \theta \) portrays the criteria. The SC’s LF \( f(x, y, \theta) \) can be expressed as

\[
    f(x, y, \theta) = \frac{1}{N} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{k} 1\{y_i = j\} \right\} \frac{1}{\sum_{j=1}^{k} \exp(\theta_j^T x_i)} \exp(\theta_j^T x_i),
\]

where \( 1\{y_i = j\} = \begin{cases} 0, & y_{i \in j} \\ 1, & y_{i \not\in j} \end{cases} \).

Imbalanced TSs might result in the training concentrating on classes with a great quantity of samples. The generalization capability with respect to the test data can be improved by fixing \( w \) within the softmax loss function (SLF), whereby tiny class samples (CSs) are multiplied by a large weight and large class samples are multiplied by a tiny
weight to diminish the class unbalance issue within this database, thus enhancing detection precision. The weighted SLF \( J(x, y, \theta) \) is calculated as follows:

\[
J(x, y, \theta) = \frac{1}{N} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{K} w_i 1\{y_i = j\} \right\} \frac{1}{b} \exp \left( \theta_j^T x_i \right) \sum_{j=1}^{K} \exp \left( \theta_j^T x_i \right),
\]

where \( w_i = \frac{M_i}{M_j} \) portrays the LF’s weight, \( M_i \) portrays the TSs, and \( M_j \) portrays the TS class sample count. The LF can be reduced using stochastic gradient descent (SGD) methodology.

5. Experimental Analysis

During training, the mini-batch SGD with a batch size of 16, momentum of 0.9, and weight decay of 0.0005 was employed. The “poly” learning rate (LR) policy was utilized in which the learning rate was multiplied by a power of 0.9, and the original LR was 0.001. The maximal number of epochs was 100 [30]. Table 1 exhibits the VGG16_CNN’s weight and biases.

| Name   | Filter     | Feature Map   | Weights | Biases |
|--------|------------|---------------|---------|--------|
| Conv3-64| 3 x 3 x 64 | 50 x 50 x 64  | 1728    | 65     |
| Conv3-64| 3 x 3 x 64 | 50 x 50 x 64  | 3456    | 56     |
| Conv3-128| 3 x 3 x 128| 50 x 50 x 128| 34,554 | 78     |
| Conv3-256| 3 x 3 x 256| 50 x 50 x 256| 34,579 | 88     |
| Conv3-256| 3 x 3 x 256| 50 x 50 x 256| 75,235 | 34     |
| Conv3-512| 3 x 3 x 512| 12 x 12 x 512| 57,561 | 36     |
| Conv3-512| 3 x 3 x 512| 12 x 12 x 512| 43,575 | 56     |
| Atrous  | 3 x 3 x 512| 3 x 3 x 512  | 34,591 | 78     |
| Atrous  | 3 x 3 x 512| 3 x 3 x 512  | 45,647 | 79     |
| Conv 1 x 1| 1 x 1 x 4 | 1 x 1 x 512  | 45,890 | 89     |
| Conv 1 x 1| 1 x 1 x 4 | 1 x 1 x 512  | 87,902 | 65     |

5.1. Database Explanation

Data were collected out for 233 patients with three BT types—meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). Because of the limitations of the depository’s file dimensions, the complete database was divided into four subcategories and obtained in four .zip files comprising 766 slices each. Fivefold cross-validation indices were also determined. Patients underwent SRSy according to the CyberKnife stereotactic therapy scheme. For therapy arrangements, MR images incorporating contrast-enhanced T1-weighted (T1c), T2-weighted, and fluid-attenuated (FLAIR) images were considered. Patients not meeting these criteria were omitted from the research. Target forms (planning target volumes—PTVs) were physically depicted on the MR images by board-authorized neurosurgeons or radiation oncologists.

5.2. Execution Metrics

In this study, accuracy, precision, recall, and F1-score were chosen as the evaluation criteria. The proffered VGG16_CNN was correlated with three conventional methodologies (moU-Net, DSNet, and U-Net) centered upon the above criteria.

Accuracy indicates the capability of the comprehensive prognosis generated by the method. The true positive (TP) and true negative (TN) depict the ability to determine the
existence or lack of MtS. The false positive (FP) and false negative (FN) depict cases of false prognosis by the employed method.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.
\] (13)

The precision rate is the proportion of positive samples to total samples with MtS.

\[
\text{Precision} = \frac{TP}{TP + FP}.
\] (14)

Recall reflects the capability to precisely identify MtS within the database; the susceptibility computation in no way considers indefinite test outcomes since the test can be reiterated and indefinite samples are omitted from assessment.

\[
\text{Recall} = \frac{TP}{TP + FN}.
\] (15)

The F1-score is used to define the prognosis, which is expressed as the weighted mean of precision and recall. A value of 1 defines good performance, whereas a value of 0 defines poor performance. F1-score in no way takes into account the TNs; it is computed as follows:

\[
\text{F1 - Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\] (16)

Figure 5 exhibits the confusion matrix for features using a classifier testing model, in which the rows represent the predicted output, and columns represent the real data. The diagonal violet, blue, and pink colors represent tested networks that were correctly and incorrectly classified. The column on the right represents every predicted class, and the row at the bottom represents the performance of every actual class.

![Confusion Matrix](image)

**Figure 5.** Confusion matrix.

Figure 6 illustrates the precision–recall curve, in which the X-axis exhibits the recall value, and the Y-axis exhibits the precision value. During this procedure, the proffered methodology attained an AP of 0.9785 for meningioma, 0.9936 for glioma, and 0.9924 for pituitary tumors, exhibiting the fine precision and recall of VGG16_CNN.

![Precision-Recall Curve](image)

**Figure 6.** Precision–recall curve correlation.

Figure 7 illustrates the precision–recall curve, in which the X-axis exhibits the FP rate, and the Y-axis exhibits the TP rate. During this procedure, the proffered methodology attained an AP of 0.9919 for meningioma, 0.9954 for glioma, and 0.9964 for pituitary tumors, exhibiting the fine precision and recall VGG16_CNN.
methodology attained an AP of 0.9785 for meningioma, 0.9936 for glioma, and 0.9924 for pituitary tumors, exhibiting the fine precision and recall of VGG16_CNN.

![Precision-Recall Curve](image)

**Figure 6.** Precision–recall curve correlation.

Figure 7 illustrates the precision–recall curve, in which the X-axis exhibits the FP rate, and the Y-axis exhibits the TP rate. During this procedure, the proffered methodology attained an AP of 0.9919 for meningioma, 0.9954 for glioma, and 0.9964 for pituitary tumors, exhibiting the fine precision and recall VGG16_CNN.

![Receiver Operating Characteristic Curve](image)

**Figure 7.** ROC curve correlation.

Table 2 presents the accuracy of existing techniques in comparison to the proposed method. Figure 8 exhibits an accuracy comparison of moU-Net, DSNet, U-Net, and the proffered VGG16_CNN methodology, in which the X-axis exhibits the quantity of epochs employed for assessment, and the Y-axis exhibits the accuracy values attained as a percentage. The moU-Net, DSNet, and U-Net methodologies attained accuracies of 92.02%, 94.3%, and 93.04%, respectively. The proffered VGG16_CNN methodology attained 93.74%...
accuracy, representing an improvement of 1.76%, 1.54%, and 0.7% compared to moU-Net, DSNet, and U-Net, respectively.

Table 2. Comparison of accuracy over various epochs between existing techniques and the proposed methodology.

| Epochs | 10   | 20   | 30   | 45   | 65   |
|--------|------|------|------|------|------|
| moU-Net   | 88.4 | 89.3 | 90.2 | 91.1 | 92.02 |
| DSNet     | 87.2 | 88.2 | 91.4 | 93.3 | 94.3 |
| U-Net     | 88.1 | 89.3 | 90.1 | 92.2 | 93.04 |
| VGG16_CNN | 90.2 | 91.3 | 93.2 | 94.3 | 95.53 |

Figure 8. Accuracy comparison of existing techniques and the proposed method.

Table 3 presents the precision value of existing techniques and the proposed method. Figure 9 exhibits a precision comparison of moU-Net, DSNet, U-Net, and the proffered VGG16_CNN methodology, in which the X-axis exhibits the number of epochs employed for assessment, and the Y-axis exhibits the precision values attained as a percentage. The moU-Net, DSNet, and U-Net methodologies attained a precision of 79.4%, 84.72%, and 86.42%, respectively. The proffered VGG16_CNN methodology attained a precision of 92%, representing an increase of 13.4%, 8.72%, and 6.42% compared to moU-Net, DSNet, and U-Net, respectively.

Table 3. Comparison of precision over various epochs between existing techniques and proposed methodology.

| Epochs | 10   | 20   | 30   | 45   | 65   |
|--------|------|------|------|------|------|
| moU-Net   | 73.4 | 74.3 | 75.2 | 77.1 | 79.4 |
| DSNet     | 77.2 | 78.2 | 81.4 | 83.3 | 84.72|
| U-Net     | 80.1 | 82.3 | 82.1 | 84.2 | 86.42|
| VGG16_CNN | 89.2 | 90.3 | 93.2 | 94.3 | 95.94|
Table 4 presents the recall of existing techniques and the proposed method. Figure 10 exhibits the recall comparison of moU-Net, DSNet, U-Net, and the proffered VGG16_CNN methodology, in which the X-axis exhibits the number of epochs employed for assessment, and the Y-axis exhibits the recall values attained as a percentage. The moU-Net, DSNet, and U-Net methodologies attained a recall of 90.7%, 91.44%, and 91.58%, respectively. The proffered VGG16_CNN methodology attained a recall of 92.1%, representing an improvement of 2.6%, 1.34%, and 1.48% compared to moU-Net, DSNet, and U-Net, respectively.

Table 4. Comparison of recall over various epochs between existing techniques and proposed methodology.

| Epochs | 10  | 20  | 30  | 45  | 65  |
|--------|-----|-----|-----|-----|-----|
| moU-Net | 86.4 | 87.3 | 88.2 | 89.1 | 90.7 |
| DSNet  | 85.2 | 86.2 | 87.4 | 88.3 | 91.44|
| U-Net  | 84.1 | 86.3 | 88.1 | 89.2 | 91.58|
| VGG16_CNN | 89.2 | 90.3 | 92.2 | 93.3 | 94.44|

Table 5 presents the F1-score of the existing techniques and proposed method. Figure 11 exhibits the F1-score comparison of moU-Net, DSNet, U-Net, and the proffered VGG16_CNN methodology, in which the X-axis exhibits the number of epochs employed for assessment, and the Y-axis exhibits the F1-score values attained as a percentage. The moU-Net, DSNet, and U-Net methodologies attained an F1-score of 64.62%, 65.24%, and 66.76%, respectively. The proffered VGG16_CNN methodology attained an F1-score of 67.08%, representing an improvement of 3.04%, 2.24%, and 2.3% compared to moU-Net, DSNet, and U-Net, respectively.
Table 5. Comparison of F1-score over various epochs between existing techniques and proposed methodology.

| Epochs (epochs) | moU-Net | DSNet | U-Net | VGG16_CNN |
|----------------|---------|-------|-------|-----------|
| 10             | 60.4    | 58.2  | 62.1  | 86.2      |
| 20             | 61.3    | 60.2  | 63.3  | 87.3      |
| 30             | 62.2    | 62.4  | 64.1  | 89.2      |
| 45             | 63.1    | 63.3  | 65.2  | 90.3      |
| 65             | 64.62   | 65.24 | 66.76 | 95.12     |

Figure 10. Recall comparison of the existing techniques and proposed methodology.

Figure 11. F1-score comparison of the existing techniques and proposed method.

Table 6. Comprehensive comparison between proposed and existing methodologies.

| Criteria   | moU-Net | DSNet | U-Net | VGG16_CNN |
|------------|---------|-------|-------|-----------|
| Accuracy   | 92.02   | 94.3  | 93.04 | 95.53     |
| Precision  | 79.4    | 84.72 | 86.42 | 95.94     |
| Recall     | 90.7    | 91.44 | 91.58 | 94.44     |
| F1-score   | 64.62   | 65.24 | 66.76 | 95.12     |

Figure 13 depicts the overall flow of the proposed methodology at various stages of the brain metastasis approach.
Table 6 presents the overall comparison of existing techniques and the proposed VGG16_CNN method, clearly revealing that the latter outperformed existing techniques in terms of the highest classified output. As shown in Figure 12, the training of features was performed to create a new feature array for training through the VGG16_CNN network.

Table 6. Comprehensive comparison between proposed and existing methodologies.

| Criteria     | moU-Net | DSNet | U-Net | VGG16_CNN |
|--------------|---------|-------|-------|-----------|
| Accuracy (%) | 92.02   | 94.3  | 93.04 | 95.53     |
| Precision (%)| 79.4    | 84.72 | 86.42 | 95.94     |
| Recall (%)   | 90.7    | 91.44 | 91.58 | 94.44     |
| F1-score (%) | 64.62   | 65.24 | 66.76 | 95.12     |

Figure 12. Overall comparison of existing techniques and the proposed method.

Figure 13 depicts the overall flow of the proposed methodology at various stages of the brain metastasis approach.

Many brain imaging tools are available to cognitive neuroscientists, including positron emission tomography (PET), near-infrared spectroscopy (NIRS), magnetoencephalography (MEG), electroencephalography (EEG), and functional magnetic resonance imaging (fMRI). The steps involved in image preprocessing are reading, resizing, removing noise (denoising), segmentation, and morphology (smoothing edges), whereas image segmentation is a method of dividing a digital image into subgroups called image segments, thereby reducing the complexity of the image and enabling further processing or analysis of each image segment. For ABMS identification and segmentation, standard machine learning methodologies such as template matching, support vector machine, and AdaBoost are implemented. The moU-Net performs well for the segmentation of glioblastoma and low-rank glioma. The U-Net is marked by an encoder that outputs lower-grade portrayals of the input data (ID) and is associated with a decoder that rebuilds the correlated label map with skip connections between intermediate phases of the two units. The prognosis of patients
with BMs is notably poor and ruins quality of life. The median survival of patients with BMs does not exceed 5 months in spite of the present therapeutic possibilities. However, the description of perfectly identified predictive subsets remains vital for the selection of a customized treatment plan. It remains significant to detect subsets of patients with suitable predictive attributes, which can then lead to the development of therapy targeting survival and an improved quality of life.

![Image of various stages of pre-processing](image)

**Figure 13.** Various stages of the pre-processing (masking), segmentation, and output image of brain analysis.

### 6. Conclusions

In this study, a DL-related classification method called VGG16_CNN was established for differentiating BMs by employing standard MR images. This method can be employed clinically by an expert radiologist to differentiate BMs, assisting in the classification of
brain MRIs into meningiomas, gliomas, and pituitary tumors. Less extensive hardware specifications are needed, and large images (256 × 256) can be processed in an appropriate time. Furthermore, the VGG16_CNN classifier exhibits finer outcomes with respect to conventional classifiers such as moU-Net, DSNet, and U-Net. It can be noted that the proffered VGG16_CNN attained 93.74% accuracy, 92% precision, 92.1% recall, and 67.08% F1-score. Hence, the novel loss function can be computed by employing the weighted soft-max function (WSF) for enhancing minute MtS diagnosis and for calibrating susceptibility and particularity.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The author declares no conflict of interest.

**References**

1. Tabouret, E.; Chinot, O.; Mettellus, P.; Tallet, A.; Viens, P.; Goncalves, A. Recent Trends in Epidemiology of Brain Metastases: An Overview. *Anticancer Res.* 2012, 32, 4655–4662. [PubMed]

2. Steinmann, D.; Schäfer, C.; van Oorschot, B.; Wypior, H.J.; Bruns, F.; Bölling, T.; Sehlen, S.; Hagg, J.; Bayerl, A.; Geinitz, H.; et al. Effects of Radiotherapy for Brain Metastases on Quality of Life (QoL). *Strahlenther. Onkol.* 2009, 185, 190–197. [CrossRef] [PubMed]

3. Le Rhun, E.; Guckenberger, M.; Smits, M.; Dummer, R.; Bachelor, T.; Sahm, F.; Galldiks, N.; de Azambuja, E.; Berghoff, A.S.; Metellus, P.; et al. EANO-ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-up of Patients with Brain Metastasis from Solid Tumours. *Ann. Oncol.* 2021, 32, 1332–1347. [CrossRef] [PubMed]

4. Chang, E.L.; Wefel, J.S.; Hess, K.R.; Allen, P.K.; Lang, F.F.; Kornguth, D.G.; Arbuckle, R.B.; Swint, J.M.; Shiu, A.S.; Maor, M.H.; et al. Neurocognition in Patients with Brain Metastases Treated with Radiosurgery or Radiosurgery Plus Whole-brain Irradiation: A Randomised Controlled Trial. *Lancet Oncol.* 2009, 10, 1037–1044. [CrossRef]

5. Kocher, M.; Wittig, A.; Piroth, M.D.; Treuer, H.; Seegenschmiedt, H.; Ruge, M.; Grosu, A.-L.; Guckenberger, M. Stereotactic Radiotherapy for Treatment of Brain Metastases. *Strahlenther. Onkol.* 2014, 190, 521–532. [CrossRef]

6. Brown, P.D.; Jaeckle, K.; Ballman, K.V.; Farace, E.; Cerhan, J.H.; Anderson, S.K.; Carrero, X.W.; Barker, F.G.; Deming, R.; Burri, S.H.; et al. Effect of Radiosurgery Alone vs. Radiosurgery with Whole Brain Radiation Therapy on Cognitive Function in Patients with 1 to 3 Brain Metastases: A Randomized Clinical Trial. *JAMA* 2016, 316, 401–409. [CrossRef]

7. Sperduto, P.W.; Mesko, S.; Li, J.; Cagney, D.; Aizer, A.; Lin, N.U.; Neshit, E.; Kruser, T.J.; Chan, J.; Braunstein, S.; et al. Beyond an Updated Graded Prognostic Assessment (breast GPA): A Prognostic Index and Trends in Treatment and Survival in Breast Cancer Brain Metastases from 1985 to Today. *Int. J. Radiat. Oncol. Biol. Phys.* 2020, 107, 334–343. [CrossRef] [PubMed]

8. Kocher, M.; Ruge, M.I.; Galldiks, N.; Lohmann, P. Applications of Radiomics and Machine Learning for Radiotherapy of Malignant Brain Tumors. *Strahlenther. Onkol.* 2020, 196, 856–867. [CrossRef] [PubMed]

9. Razzak, M.I.; Naz, S.; Zaib, A. Deep learning for medical image processing: Overview, challenges and the future. *Classif. BioApps* 2018, 3, 323–350.

10. Zhang, Y.; Li, H.; Du, J.; Qin, J.; Wang, T.; Chen, Y.; Liu, B.; Gao, W.; Ma, G.; Lei, B. 3D Multi-attention Guided Multi-task Learning Network for Automatic Gastric Tumor Segmentation and Lymph Node Classification. *IEEE Trans. Med. Imaging* 2021, 40, 1618–1631. [CrossRef]

11. Chen, J.; Jiao, J.; He, S.; Han, G.; Qin, J. Few-shot Breast Cancer Metastases Classification via Unsupervised Cell Ranking. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2019, 18, 1914–1923. [CrossRef] [PubMed]

12. Wang, L.; Song, T.; Katayama, T.; Jiang, X.; Shimamoto, T.; Leu, J.S. Deep Regional Metastases Segmentation for Patient-Level Lymph Node Status Classification. *IEEE Access* 2021, 9, 129293–129302. [CrossRef]

13. Isensee, F.; Petersen, J.; Kohl, S.A.; Jäger, P.F.; Maier-Hein, K.H. nnUNet: Breaking the Spell on Successful Medical Image Segmentation. *arXiv* 2019, arXiv:190408128.

14. Kickingreder, P.; Isensee, F.; Tursunova, I.; Petersen, J.; Neuberger, U.; Bonekamp, D.; Brugnara, G.; Schell, M.; Kessler, T.; Foltyn, M.; et al. Automated Quantitative Tumour Response Assessment of MRI in Neuro-oncology with Artificial Neural Networks: A multicentre, retrospective study. *Lancet Oncol.* 2019, 20, 728–740. [CrossRef]

15. Xue, J.; Wang, B.; Ming, Y.; Liu, X.; Jiang, Z.; Wang, C.; Liu, X.; Chen, L.; Qu, J.; Xu, S.; et al. Deep-Learning-Based Detection and Segmentation-Assisted Management on Brain Metastases. *Neuro-Oncology* 2019, 22, 505–514. [CrossRef]

16. McBee, M.P.; Awan, O.A.; Colucci, A.T.; Ghobadi, C.W.; Kadom, N.; Kansagra, A.P.; Tridandapani, S.; Auffermann, W.F. Deep Learning in Radiology. *Acad. Radiol.* 2018, 25, 1472–1480. [CrossRef]

17. Samani, Z.R.; Parker, D.; Wolf, R.; Hodges, W.; Brem, S.; Verma, R. Distinct Tumor Signatures using Deep Learning-based Characterization of the Peritumoral Microenvironment in Glioblastomas and Brain Metastases. *Sci. Rep.* 2021, 11, 14469. [CrossRef]
18. Dong, F.; Li, Q.; Jiang, B.; Zhu, X.; Zeng, Q.; Huang, P.; Chen, S.; Zhang, M. Differentiation of supratentorial single brain metastasis and glioblastoma by using peri-enhancing oedema region-derived radiomic features and multiple classifiers. *Eur. Radiol.* 2020, 30, 3015–3022. [CrossRef]

19. Shin, I.; Kim, H.; Ahn, S.; Sohn, B.; Bae, S.; Park, J.; Kim, H.; Lee, S.-K. Development and Validation of a Deep Learning–Based Model to Distinguish Glioblastoma from Solitary Brain Metastasis Using Conventional MR Images. *Am. J. Neuroradiol.* 2021, 42, 838–844. [CrossRef]

20. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition in Computer Vision and Pattern Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

21. Huang, G.; Liu, Z.; der Maaten, L.V.; Weinberger, K.Q. Densely Connected Convolutional Networks in Computer Vision and Pattern Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.

22. Li, J.; Ng, W.W.Y.; Tian, X.; Kwong, S.; Wang, H. Weighted Multi-deep Ranking Supervised Hashing for Efficient Image Retrieval. *Int. J. Mach. Learn. Cybern.* 2020, 11, 883–897. [CrossRef]

23. Yu, X.; Fan, J.; Chen, J.; Zhang, P.; Zhou, Y.; Han, L. NestNet: A multiscale convolutional neural network for remote sensing image change detection. *Int. J. Remote Sens.* 2021, 42, 4898–4921. [CrossRef]

24. Khaleghian, S.; Ullah, H.; Krämer, T.; Hughes, N.; Eltoft, T.; Marinoni, A. Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks. *Remote Sens.* 2021, 13, 1734. [CrossRef]

25. Losch, M. Detection and Segmentation of Brain Metastases with Deep Convolutional Networks. Master’s Thesis, KTH, Computer Vision and Active Perception, CVAP. 2015. Available online: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-173519 (accessed on 15 June 2022).

26. Zhong, L.; Meng, Q.; Chen, Y.; Du, L.; Wu, P. A laminar augmented cascading flexible neural forest model for classification of cancer subtypes based on gene expression data. *BMC Bioinform.* 2021, 22, 1–17.

27. Pang, S.; Fan, M.; Wang, X.; Wang, J.; Song, T.; Wang, X.; Cheng, X. VGG16-T: A novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by CT images. *Int. J. Comput. Intell. Syst.* 2020, 13, 771. [CrossRef]

28. Renjith, V.S.; Jose, P.S.H. Efficacy of Deep Learning Approach for Automated Melanoma Detection. In Proceedings of the 2021 International Conference on Decision Aid Sciences and Application (DASA), Virtual, 7–8 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 471–478.

29. Cheng, G.; Kadry, S.; Meqdad, M.N.; Crespo, R.G. CNN supported framework for automatic extraction and evaluation of dermoscopy images. *J. Supercomput.* 2022, 78, 17114–17131. [CrossRef]

30. Arcos-Garcia, A.; Alvarez-Garcia, J.A.; Soria-Morillo, L.M. Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods. *Neural Netw.* 2018, 99, 158–165. [CrossRef] [PubMed]