



Handling the predictive uncertainty of convolutional neural network in medical image analysis: a review

Yasanthi Malika Hirimutugoda^{1^}, Thusari P. Silva¹, Nimalka M. Wagarachchi²

¹Faculty of Information Technology, University of Moratuwa, Moratuwa, Sri Lanka; ²Faculty of Engineering, University of Ruhuna, Galle, Sri Lanka

Contributions: (I) Conception and design: All authors; (II) Administrative support: TP Silva, NM Wagarachchi; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: None; (V) Data analysis and interpretation: None; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yasanthi Malika Hirimutugoda, MSc. Faculty of Information Technology, University of Moratuwa, Bandaranayake Mawatha, Moratuwa 10400, Sri Lanka. Email: yasanthimalika@yahoo.com.

Abstract: Certainty is a significant part of disease detection, involving various kinds of imaging and machine learning (ML) methodologies. More precisely than other ML methods, a convolutional neural network (CNN) can classify images. As its parameters are deterministic, it cannot indicate the level of uncertainty in its predictions. Predictions made by predetermined CNNs may yield inaccurate findings, and there is no evaluation of confidence in these results. These outcomes may have harmful effects and lack trustworthiness. Uncertainty quantification (UQ) is critical to evaluating confidence in prediction. The noise, illumination, segmentation, and edge issues common to medical images also impact pre-trained CNN algorithms and lead to uncertain outcomes. This review aims to investigate the main inherent uncertainty issue in CNN and what form of UQ method can be applied with CNN to the task of medical image classification. This research proposes a novel approach by combining the superior properties of the Bayesian approach and fusion methods to reduce the uncertainty in CNN models. This study concludes that, despite a number of unresolved technical and scientific issues, various types of fusion approaches have improved the clinical validity for diagnosing and analytical purposes, and it is a field of study that has the capacity to grow significantly in the years to come.

Keywords: Convolutional neural network (CNN); machine learning (ML); uncertainty; Bayesian; fusion

Received: 01 May 2023; Accepted: 22 September 2023; Published online: 16 October 2023.

doi: 10.21037/jmai-23-40

View this article at: <https://dx.doi.org/10.21037/jmai-23-40>

Introduction

Certainty is an essential part of medical image detection systems, in conjunction with accurate interpretation. The complexity of the images, differences in explanation approaches, subjectivity, accuracy, and throughput are prone to errors in medical diagnosis. The aim of automated medical image analysis is to help radiologists and clinicians improve the efficiency of diagnostic and treatment processes. But convolutional neural networks

(CNNs) have the inherent capability to learn complicated characteristics directly from the source data, which is a significant advantage in the medical domain (1). Computer-assisted detection systems (CADx) heavily rely on machine learning (ML) for various image analysis tasks such as tumor segmentation, cancer detection, image-directed treatment, pathological brain detection, medical image annotation, Alzheimer's disease detection, diabetic retinopathy, and image capture. X-ray, computed tomography (CT), magnetic resonance imaging (MRI), positron emission

[^] ORCID: 0000-0003-1687-8195.

tomography (PET), PET/CT hybrid, and three-dimensional ultrasound computed tomography (3D USCT) are examples of various hybrid modalities. Key parameters, including “accuracy”, “F-measure”, “precision”, “recall”, “sensitivity”, and “specificity” must perform well in medical imaging systems (2).

Several researchers have introduced different kinds of pre-trained deep convolutional neural networks (DCNNs) such as AlexNet (3), ResNet (4), VGG (5), and GoogleNet (6). However, current DCNN models still do not completely handle ongoing segmentation and classification issues. The noise and illumination issues that are common to medical images also have an impact on these pre-trained CNN algorithms. Applying image pre-processing techniques can reduce noise, improve performance, and increase the predictability of CNN results (7).

It is important to evaluate uncertainty in CNNs in order to guarantee decisions. Nonetheless, despite the fact that DL algorithms typically outperform conventional methods, their mapping process is based on blind assumptions that frequently do not accurately represent real-world cases. Selecting factors that have an impact on producing a particular result requires careful consideration of uncertainty and interpretability. There is no built-in way to provide explanations for a specific result in a CNN-based decision support system. DL requires a massive amount of data for data training, with millions of parameters. It is a significant obstacle in many disciplines, such as the medical domain. Pre-trained CNN models from different domains and data augmentation are some ways to address the aforementioned problems, and they have produced amazing results. However, there is currently a need for research and development in this area, particularly when quantification uncertainty and interpretability are both improved (1).

This review aims to present the main inherent issues in CNN and what techniques can be used to solve these issues, making it easier for new researchers to understand the uncertainty in CNN. A new hybrid model using feature fusion techniques and Bayesian algorithms is proposed for classifying medical images to address above mentioned key issues with CNN. By enabling readers to learn more about current advancements in the medical image analysis field, this review will strengthen subsequent deep CNN innovations. Researchers would be free to choose the best course of action to select the more reliable alternatives in this field. Our contributions are outlined as follows: This review paper is prearranged in the following order: “*Introduction*” section includes future research gaps in CNN and the uncertainty of

CNN, the significance of uncertainty quantification (UQ) in medical image analysis, and forms of UQ in CNN. “*Literature review*” section includes evaluation of deep learning and CNN. “*Analysis of knowledge for UQ*” section represents analysis of various methods with CNN for UQ. Finally, “*Discussion*” and “*Conclusions*” sections have been arranged.

Future research gaps of CNN and uncertainty of CNN

Despite CNN models having remarkable performance and ability for natural feature learning, CNN does not provide naturally accurate uncertainty estimates for their segmentation process. The need for uncertainty assessment in deep CNN-based medical imaging segmentation techniques is, therefore, essential (8). Despite CNN outperforming other ML approaches in terms of accuracy for image classification, its deterministic parameters prevent it from being able to provide any sort of measure of uncertainty in its prediction. Additionally, predefined CNN predictions could lead to erroneous results that, in the absence of a confidence estimate, could have negative implications. UQ is necessary to measure the confidence of predictions (9).

Thiagarajan *et al.* have shown that impossibility to determine how accurate CNN-based classification algorithms for histopathology images and unbalanced data might cause overfitting in CNNs. That has demonstrated how automatic regularization and UQ allows Bayesian-CNN to get over the above drawbacks (10). Since medical imaging technologies function as “black box” devices, medical experts may dismiss them because they cannot comprehend the calculations or how findings are created. To demonstrate to medical professionals how computations are done and how parameters can affect a computational outcome, visualization techniques like “Medical Imaging Interaction Toolkit” are necessary in this situation. With this program, users can manipulate images and see the results of computations as they are done. The issue of using these methods with ML in medical imaging still exists for further research (11).

Hussain *et al.* present a deep CNN model for brain tumor segmentation. After segmentation, some false positives may show up because of the frequency underlying the skull fraction. To minimize small false positives at the segmented image’s edges, the opening and closing morphological operators are used, which apply erosion and dilation operations consecutively. AI is a “black box”, meaning that while its input and output can be identified, its internal representations are not fully understood. It also

has an impact on medical image analysis. The noise and illumination issues present in medical images also have an impact on these technologies (8).

The significance of UQ in medical image analysis

A lack of confidence in each outcome that a ML system generates is “uncertainty”. It is difficult to create an algorithm that is 100 percent certain, so we need to understand what creates uncertainty, how to define it, and how to eliminate it. ML is now widely used in medical applications to predict diseases, arrange treatment, and evaluate performance. Due to the lack of necessary trust among practitioners, the application of this technology in a clinical environment is still very uncommon (11).

Estimating uncertainty is essential since deep learning (DL) tends to produce overly confident predictions. Overconfidence and erroneous predictions could be dangerous in important use cases like healthcare. In addition to providing findings, the model would also provide details regarding their level of uncertainty and whether or not it is low enough to warrant confidence in the output. The algorithm may request further data or input from a person to handle the decision-making process before providing an output with a high level of uncertainty. With these inputs, we want our model to be able to express a given quantity that is highly uncertain or lowly confident (12).

Sources of uncertainty and UQ in CNN

Uncertainty comes in two primary forms. They are model uncertainty or knowledge uncertainty [epistemic uncertainty (EU)] and data uncertainty [aleatoric uncertainty (AU)]. Predictive uncertainty (PU), or the level of confidence in a prediction, can then be created using AU and EU. AU results from noise and class overlap in the data and EU results from the inconsistency between the training and test sets of data.

AU occurs due to complexity, multi-modality, illuminations, corruptions, and unseen noise in the data. Since data uncertainty is a characteristic of the fundamental distribution that produced the data rather than a feature of the model, it is impossible to reduce it. Given specific conditions, probabilistic classification and regression models will automatically capture estimates of AU as a result of maximum likelihood training. However, it is more challenging to capture estimates of EU. Epistemic or model uncertainty refers to uncertainty in model predictions

resulting from model ignorance or a lack of comprehension of the current input on which the model is building a forecast. Unlike data uncertainty, model uncertainty occurs due to the absence of features in the data. As knowledge uncertainty is a property of the model, it can be decreased by giving more information in the form of training data to the model (13). EU and AU are the two components that make up PU. PU can be expressed as the total of these two components: $EU + AU = PU$.

UQ in CNN is a crucial component of DL, and it is especially significant in the analysis of medical data since the diagnosis results from various approaches will directly affect actual life situations. In order to increase the effectiveness of ML and produce greater confidence in the results, accurate uncertainty estimations are necessary (14). Even if few models have been proposed for different UQ methods of CNN, it remains an unsolved problem yet.

Although CNN is more accurate than other ML algorithms at classifying images, its parameters are deterministic, so it is unable to provide any indication of the degree of uncertainty in its predictions. Furthermore, predictions made by deterministic CNNs may yield inaccurate findings, and as there is no evaluation of confidence in these results, these outcomes may have unfavorable effects (9). Model performance can be achieved by examining the contribution of each pixel to predict uncertainty and the drop-in prediction uncertainty following the removal of a noisy pixel from the input image.

Literature review

Evaluation of deep learning and CNN

The capability of the machine to extract features or knowledge from input and to learn from experience is known as ML. This technology allows computers to draw complicated conclusions based on the relationships in the data (14). Statistical analysis would be impractical for handling huge and complicated datasets, but ML can. It can construct predictive models without using predetermined coding rules (15). ML algorithms fall into three types: reinforcement learning, unsupervised learning, and supervised learning (1).

Feature engineering is the key component of ML algorithms. ML algorithms significantly rely on feature engineering. It is a method of applying statistical or ML techniques to transform unprocessed information into desired characteristics. To achieve the best results, the feature

selection process takes a lot of time and energy and features must retrieve pertinent information from massive amounts of different data. However, feature engineering is a difficult issue for ML algorithms in unstructured domains like computer vision, image processing, video and audio analysis, and natural language processing. To address these issues, a specific subfield of ML called DL, which is founded on the theory of artificial neural networks (ANNs) emerged (1).

The amount of hidden layers is the main distinction between DL and conventional ANNs (14). To investigate more complicated nonlinear patterns and discover meaningful links within the input dataset, DNNs employ many layers. By decreasing or even eliminating the requirement for feature engineering, DNN learns and creates unique characteristics from each subsequent hidden layer of neurons. Due to the last of these factors, DL techniques frequently do better than ML algorithms (16), innovating the field with successful performance and adaptability to input noise and inconsistency in a variety of tasks (17).

DNN architecture may be categorized into three main groups. They are feed-forward neural networks (also known as multi-layer perceptron, MLP), CNNs, and recurrent neural networks (RNN) (18). CNN is one of the most well-known and significant developments in computer vision. DL algorithms are influenced by the neurobiological structure of the visual cortex and they are initially supported in image analysis (19). Artificial neurons in the convolutional layer compute a spatial convolution, capturing characteristics from minuscule sections of the input images while training.

Analysis of knowledge for UQ

Hybrid CNN architecture for UQ

This chapter provides a general overview of the field of UQ for discriminative parametric classification and regression models. Both epistemic and aleatoric uncertainties for deep CNNs have been researched recently. Researchers have developed a range of approximation techniques using the fundamental CNN design to reduce uncertainty (20).

CNN with Bayesian approach for UQ

This section examines the preliminary results of Bayesian neural networks (BNNs) and uncertainty varieties. Although the typical CNN approaches have been successful in resolving a number of real-world issues, they are unable to reveal any information regarding the accuracy of their

forecasts. This issue can be solved by translating the model parameters using Bayesian deep learning (BDL). BDLs are very strong at overcoming uncertainty and over-fitting issues and can be trained on both minor and massive datasets (21). UQ models with unreliable predictions are necessary to establish the validity of segmentations. This model demonstrated that the Bayesian approach generates more believable confidence intervals on the segmentation than the Monte Carlo Dropout for CT images.

Future research will involve expanding our BCNN to semantic segmentation, separating AU and EU, and fusing our findings with other UQ methods like deep ensembles (22). BCNN accomplishes improved segmentation accuracy than MCDNs and outperforms MCDNs in recent uncertainty metrics. Future research should focus on creating metrics that gauge how well Bayesian models function based on segmentation output and uncertainty estimates (23). *Table 1* lists some BCNN models with their model performance and drawbacks for future developments of BCNN. This analysis shows that BCNN alone can't adequately solve the uncertainty issue in medical imaging.

CNN with other methods for UQ

Theoretically, a Bayesian approach allows for the effective execution of uncertainty reasoning. In addition to that, all key information is presented by probability distributions and various sources of information can be combined using the Bayesian formula. But still, this offers significant computing difficulties, particularly given the sophisticated models that have emerged in DL. The classical approximate inference approaches [such as Markov chain Monte Carlo (MCMC), Laplace approximation, and variational inference] as well as the more modern BNNs and Monte Carlo dropout have all been thoroughly addressed. Also, researchers have addressed how to quantify the sources of uncertainty and provided pertinent links to open-source implementations that are stored in GitHub repositories (29).

Stochastic variational gradient descent (SVGD) has proposed to perform approximate Bayesian inference on uncertain CNN parameters (30), adaptive MCMC method with CNN (31), Bayesian deep convolutional encoder-decoder networks for substitute modeling and UQ (30), deep ensembles with CNN for robust UQ (32). To investigate the location of test data, test-time augmentation for AU estimates has been presented. They compared and combined model (epistemic) and input-based (aleatoric) uncertainty sources to examine several forms of uncertainties for CNN-based medical

Table 1 Uncertainty quantification using BCNN model with strengths and drawbacks

Study	Strength	Drawbacks and future developments
Diagnosis of blood cancer using microscopic images of blood samples (24)	Accuracy: 94%, offers helpful information regarding prediction uncertainty	CNN can be improved with more modern visualization techniques
Classifiers for breast histopathology image (10)	Novel stochastic adaptive activation enabled BCNN. It can reduce the false negative and false positive predictions by 3% as compared to BCNN	These results need to be confirmed with bigger data sets
Guidance in robotic knee arthroscopy using 4D ultrasound images (25)	Improved segmentation accuracy when training with MRI and US-based data sets. It has improved the coefficient increase by 6% and 8%, respectively	To improve the effectiveness of label propagation, implementing automatic feature-based MRI-US registrations will be considered
Tuberculosis identification using chest X-ray images (23)	Accuracy: 96.42% and 86.46% for both datasets (i.e., Montgomery and Shenzhen)	To remove uncertainty from the recognition of natural images, BCNN models will be used
Skin cancer classification (26)	TWD theory, accuracy, F1-score, and AUC are 90.96%, 91.00%, and 0.97	Noise detection evaluation and improvement of the effectiveness of the model's decisions are needed, gap decision theory and a non-probabilistic decision theory, can be applied to enhance failure robustness
COVID-19 diagnosis using CT chest images (27)	Bayesian optimization is used to choose hyperparameters for each ML method. The diagnosis made using these hyperparameters, the DenseNet201 model, and the SVM algorithm gets the highest success rate of 96.29%	Combining the features from various CNN models is intended to provide more robust features. Feature selection techniques will be used to select the best features from this large set
Diabetic retinopathy classification using CT images (28)	Binary class classification accuracy: 92% (less confident) and 81% (more confident), multi-class model accuracy: 92% (more confident)	Future research should continue to optimize these models for CNN and BCNN

BCNN, Bayesian convolutional neural network; COVID-19, coronavirus disease 2019; CT, computed tomography; MRI, magnetic resonance imaging; US, ultrasound; TWD, Three-Way Decision; AUC, area under the curve; ML, machine learning; SVM, support vector machine; CNN, convolutional neural network.

image segmentation. They developed an AU estimator for medical images based on test-time augmentation that considers the impact of both noise and spatial changes. However, its effectiveness for noise removal and medical image segmentation has not been thoroughly proven (33).

These various sorts of uncertainty were examined in this work for tasks involving the segmentation of 2D and 3D medical images at the pixel and structural levels using CNN. In addition, they suggested a test-time augmentation-based AU to examine how various input image alterations affect the output of segmentation. Although test-time augmentation has been utilized to increase segmentation performance, it has not been developed within a coherent mathematical framework. As a result, a theoretical definition of test-time augmentation has been proposed by them, in which the probability of the prediction is calculated using

Monte Carlo simulation using earlier distributions of parameters in an image-capturing model that includes noise and image modifications (34). Additionally, to overcome EU and AU, different types of integrated models using CNN are analyzed. In conclusion, taking into consideration the impact of both noise and spatial changes in images is crucial for developing AU and EU estimation for medical images (34).

Fusion methods with DL

Why are fusion techniques needed in medical image analysis?

Medical imaging diagnosis highly depends on the combination of clinical data from several sources in order to precisely interpret the imaging data and make a reliable diagnosis (35). Information fusion is initiated from data

fusion. It can also be termed multi-sensor information fusion, feature fusion for combining different features (36). Data fusion is the combining of data from numerous sensors (either of the same or different types), whereas information fusion is the combining of data and information from sensors, human reports, and databases, in addition to a wide range of contextual data (36). Medical image fusion is the process of registering and integrating several images from one or more imaging techniques (such as X-ray, CT scan) to enhance there are three types of fusion models: early fusion, late fusion, and joint fusion. Early fusion is the combining of features or feature representations before feeding them into a model. The prediction probabilities of various single modality models are combined in late fusion, often referred to as decision-level fusion, to produce an outcome. Joint fusion takes learned feature representations from each modality and appends them as inputs to another model (35). Multi-modal medical Image Fusion is the technique of combining several medical images from various modalities into a single fused image. The primary goal of medical image fusion is to gather a large amount of acceptable information (i.e., features) in order to increase its quality and make it more informative for boosting clinical care for improved diagnosis and straightforward evaluation of medicine-related uncertainties (37). Several researches have shown that brain tumor and tissue anatomy can be successfully analyzed by removing noise (38) and extracting features (39) from CT and MRI images using fusion methods.

Image fusion method with DL

This section examines the preliminary results of fusion methods and medical image analysis with CNN. Although the typical CNN approaches have been successful in resolving a number of real-world issues, they are unable to reveal any information regarding the accuracy of their predictions. *Table 2* indicates that the fusion method can provide high accuracy than Bayesian CNN. But a trustworthy model should have features of both accuracy and certainty.

Visualizing CNNs and evaluating trust in a model

Gradient-weighted class activation mapping plus (Grad-CAM++) has been utilized to improve the features' intuitiveness and the DL model's interpretability (41). This model suggests the use of Grad-CAM++ (45). Grad-CAM makes advantage of the gradient of the classification

result in relation to the network's convolutional features to decide which areas of the picture are most crucial for categorization (20). Hamza *et al.* proposed that visualization is carried out utilizing a Grad-CAM that draws attention to the area of the image that is affected. For the testing approach, three publicly accessible COVID-19 datasets are employed to be produced enhanced levels of accuracy of 98.8%, 97.9%, and 99.4% (46).

In this study, researchers presented Ablation-CAM, a class-discriminative localization technique that can produce gradient-free visual explanations. They illustrate scenarios where Grad-CAM gives less accurate representations than Ablation-CAM. They demonstrate that Ablation-CAM has solved the drawbacks of Grad-CAM visualizations. Further research is done to demonstrate that Ablation-CAM is class unfair and that it may be used to estimate model trust. This approach, in contrast to earlier ones, is "gradient-free" (47).

Optimization algorithm

An optimization method involves determining the best values for a system's particular parameters to satisfy the system design as efficiently as possible. An optimization process refers to finding the optimal values for specific parameters of a system to accomplish the system design at the lowest cost. Numerous novel optimization techniques have been developed to improve the functioning of various systems and reduce computation costs. By adding to the various optimization techniques, the performance of the model can be enhanced even more (42). The arithmetic optimization algorithm (AOA) is developed in this study based on the actions of arithmetic operators in mathematical calculations. According to its mathematical presentation, AOA has an implementation that is simple and uncomplicated compared to the majority of the commonly used optimization algorithms, making it easy to modify to address new optimization problems. Except for the size of the population and stopping criterion, which are default parameters in all optimization methods, it doesn't need to change many other settings (48). An innovative population-based optimization method called the Aquila Optimizer (AO) is inspired by the natural behaviors of Aquila as they seek their target (49). Intelligent algorithms that simulate the principles of the human immune system are called artificial immune systems (AIS). Gradient-based optimization techniques have demonstrated their effectiveness in learning over-parameterized and sophisticated neural networks from

Table 2 Fusion methods and CNN with strengths and drawbacks

Study	Model methods	Accuracy and strengths	Further improvements and drawbacks
Categorize cases of PE using CT images (35)	Late fusion model, 3D CNN, and ElasticNet	95%	Validation on an external test set from various organizations has to be done to better understand the generalizability of this model
Classifying numerous stomach disorders using endoscopy (7)	Fusion images, ResNet101, Deep transfer learning. Various modalities	99.46%	Considering more clinical data and training a CNN model from scratch
Recognition of stomach diseases using endoscopy images (40)	Transfer learning using, dragonfly optimization, CNN models, feature fusion	99.8%	To segment polyps and ulcers, the model will be trained from scratch. The feature concatenation necessitates an increase in computing cost
COVID-19 diseases prediction using X-ray images (41)	Deep feature fusion, EfficientNetV2, multiple-way data augmentation, Grad-CAM++	99.89%	It fails to express an opinion on the COVID-19 grade. It is unable to handle datasets created by combining CT and CXR. They have made suggestions for gamma corrections and noise reduction
COVID detection using CT and X-ray image set (42)	Feature fusion, CNN, Monte Carlo dropout, Grad-CAM++	99.08% and 96.35% for CT scan and X-ray datasets, model was generally robust to noise	Utilizing multi-modal data and incorporating an attention mechanism when merging features. Combining some cutting-edge data fusion techniques, such as decision-level fusion
Breast cancer detection using mammographic images (43)	Feature fusion, ensemble learning	99.4% on the MIAS dataset, and 98.8% on the BCDR test set	Merging distinct layers of patterns with varying weights to produce high accuracy
COVID-19 detection using CT images (20)	GCN, DFF	Better performance for DFF than GCN	Extend the model for X-ray and CT combination. Investigate further CNN and GCN fuse techniques. Try using deeper GCN and see if it makes a difference in performance
Telemedicine using CT and MRI (44)	Advanced CNN, multi-modal medical image fusion	Richer textual details, sharper edging information, and greater contrast	Integrate three or more modalities simultaneously

CNN, convolutional neural network; PE, pulmonary embolism; CT, computed tomography; COVID-19, coronavirus disease 2019; MRI, magnetic resonance imaging; Grad-CAM++, gradient-weighted class activation mapping plus; MIAS, Mammogram Image Analysis Society dataset; BCDR, breast cancer digital repository; GCN, graph convolutional network; DFF, deep feature fusion; CXR, chest X-ray.

non-convex objectives (42).

Discussion

For this analysis, 49 research articles from 2018 to 2023 were examined. It concludes that, despite a number of unresolved technical and scientific issues, medical image fusion has improved the clinical validity for diagnosing and analytical purposes, and it is a field of study that has the capacity to grow significantly in the years to come.

Accuracy in medical image analysis can be increased with certainty if Fusion and Bayesian approaches are combined with CNN.

Conclusions

This study suggests a novel approach and a justification for combining the Bayesian approach, the Feature Fusion method, and the Grad-CAM visualization in order to lower the degree of uncertainty in the predictions made by CNN

models for medical image analysis and increase accuracy.

Acknowledgments

Funding: None.

Footnote

Peer Review File: Available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-40/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jmai.amegroups.com/article/view/10.21037/jmai-23-40/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Bote-Curiel L, Muñoz-Romero S, Gerrero-Curieses A, et al. Deep Learning and Big Data in Healthcare: A Double Review for Critical Beginners. *Appl Sci* 2019;9:2331.
2. Anwar SM, Majid M, Qayyum A, et al. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst* 2018;42:226.
3. Hosny KM, Kassem MA, Fouad MM. Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet. *J Digit Imaging* 2020;33:1325-34.
4. Sarwinda D, Paradisa RH, Bustamam A, et al. Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Comput Sci* 2021;179:423-31.
5. Ghosh S, Chaki A, Santosh KC. Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Phys Eng Sci Med* 2021;44:703-12.
6. Yao X, Wang X, Karaca Y, et al. Glomerulus Classification via an Improved GoogLeNet. *IEEE Access* 2020;8:176916-23.
7. Khan MA, Sarfraz MS, Alhaisoni M, et al. StomachNet: Optimal Deep Learning Features Fusion for Stomach Abnormalities Classification. *IEEE Access* 2020;8:197969-81.
8. Hussain S, Anwar SM, Majid M. Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network. *Neurocomputing* 2018;282:248-61.
9. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat Mach Intell* 2019;1:20-3.
10. Thiagarajan P, Khairnar P, Ghosh S. Explanation and Use of Uncertainty Quantified by Bayesian Neural Network Classifiers for Breast Histopathology Images. *engrXiv* 2020. Available online: <https://engrxiv.org/5xf8c/>
11. Gillmann C, Saur D, Scheuermann G. How to deal with Uncertainty in Machine Learning for Medical Imaging? In: 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREV). 24-25 October 2021; New Orleans, LA, USA. IEEE; 2021:52-8.
12. Malinin A. Uncertainty Estimation in Deep Learning with application to Spoken Language Assessment. Available online: <https://doi.org/10.17863/CAM.45912>.
13. Gal Y. Uncertainty in Deep Learning. Available online: <https://www.cs.ox.ac.uk/people/yarin.gal/website/thesis/thesis.pdf>
14. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018;19:1236-46.
15. McBee MP, Awan OA, Colucci AT, et al. Deep Learning in Radiology. *Acad Radiol* 2018;25:1472-80.
16. Faust O, Hagiwara Y, Hong TJ, et al. Deep learning for healthcare applications based on physiological signals: A review. *Comput Methods Programs Biomed* 2018;161:1-13.
17. Ribeiro Pinto J, Cardoso JS, Lourenço A. Evolution, Current Challenges, and Future Possibilities in ECG Biometrics. *IEEE Access* 2018;6:34746-76.
18. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 2018;15:20170387.
19. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230-43.
20. Wang SH, Govindaraj VV, Górriz JM, et al. Covid-19 classification by FGCNet with deep feature fusion from

- graph convolutional network and convolutional neural network. *Inf Fusion* 2021;67:208-29.
21. Abdar M, Pourpanah F, Hussain S, et al. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Inf Fusion* 2021;76:243-97.
 22. LaBonte T, Martinez C, Roberts SA. We Know Where We Don't Know: 3D Bayesian CNNs for Credible Geometric Uncertainty. *arXiv* 2020. Available online: <http://arxiv.org/abs/1910.10793>
 23. Ul Abideen Z, Ghafoor M, Munir K, et al. Uncertainty Assisted Robust Tuberculosis Identification With Bayesian Convolutional Neural Networks. *IEEE Access* 2020;8:22812-25.
 24. Billah ME, Javed F. Bayesian Convolutional Neural Network-based Models for Diagnosis of Blood Cancer. *Appl Artif Intell* 2022;36:2011688.
 25. Antico M, Sasazawa F, Takeda Y, et al. Bayesian CNN for Segmentation Uncertainty Inference on 4D Ultrasound Images of the Femoral Cartilage for Guidance in Robotic Knee Arthroscopy. *IEEE Access* 2020;8:223961-75.
 26. Abdar M, Samami M, Dehghani Mahmoodabad S, et al. Uncertainty quantification in skin cancer classification using three-way decision-based Bayesian deep learning. *Comput Biol Med* 2021;135:104418.
 27. Aslan ME, Sabanci K, Durdu A, et al. COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. *Comput Biol Med* 2022;142:105244.
 28. Ahsan MA, Qayyum A, Qadir J, et al. An Active Learning Method for Diabetic Retinopathy Classification with Uncertainty Quantification. *arXiv* 2020. Available online: <http://arxiv.org/abs/2012.13325>
 29. Barbano R, Arridge S, Jin B, et al. Uncertainty quantification in medical image synthesis. In: *Biomedical Image Synthesis and Simulation Elsevier*; 2022:601-41.
 30. Zhu Y, Zabaras N. Bayesian Deep Convolutional Encoder-Decoder Networks for Surrogate Modeling and Uncertainty Quantification. *J Comput Phys* 2018;366:415-47.
 31. Zhou Z, Tartakovsky DM. Markov Chain Monte Carlo with Neural Network Surrogates: Application to Contaminant Source Identification. *Stoch Environ Res Risk Assess* 2021;35:639-51.
 32. Sahay R, Ries D, Zollweg JD, et al. Hyperspectral Image Target Detection Using Deep Ensembles for Robust Uncertainty Quantification. In: *2021 55th Asilomar Conference on Signals, Systems, and Computers*; 31 October 2021 - 03 November 2021; Pacific Grove, CA, USA. IEEE; 2021:1715-9.
 33. Ayhan MS, Berens P. Test-time Data Augmentation for Estimation of Heteroscedastic Aleatoric Uncertainty in Deep Neural Networks. *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*, Amsterdam, The Netherlands. Available online: <https://www.cs.ox.ac.uk/people/yarin.gal/website/thesis/thesis.pdf>
 34. Wang G, Li W, Aertsen M, et al. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing (Amst)* 2019;335:34-45.
 35. Huang SC, Pareek A, Zamanian R, et al. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci Rep* 2020;10:22147.
 36. Shanshan Z. Multi-source Information Fusion Technology and Its Engineering Application. *Res Health Sci* 2020;4:408.
 37. Haribabu M, Guruviah V, Yogarajah P. Recent advancements in multimodal medical image fusion techniques for better diagnosis: an overview. *Curr Med Imaging* 2023;19:673-94.
 38. Lin H, Song Y, Wang H, et al. Multimodal Brain Image Fusion Based on Improved Rolling Guidance Filter and Wiener Filter. *Comput Math Methods Med* 2022;2022:e5691099.
 39. Ding Z, Zhou D, Nie R, et al. Brain Medical Image Fusion Based on Dual-Branch CNNs in NSST Domain. *Biomed Res Int* 2020;2020:6265708.
 40. Mohammad F, Al-Razgan M. Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification. *Sensors (Basel)* 2022;22:2801.
 41. Liu J, Sun W, Zhao X, et al. Deep feature fusion classification network (DFFCNet): Towards accurate diagnosis of COVID-19 using chest X-rays images. *Biomed Signal Process Control* 2022;76:103677.
 42. Abdar M, Salari S, Qahremani S, et al. UncertaintyFuseNet: Robust uncertainty-aware hierarchical feature fusion model with Ensemble Monte Carlo Dropout for COVID-19 detection. *Inf Fusion* 2023;90:364-81.
 43. Haq IU, Ali H, Wang HY, et al. Feature fusion and Ensemble learning-based CNN model for mammographic image classification. *J King Saud Univ Comput Inf Sci* 2022;34:3310-8.
 44. Liu Y, Yan B, Zhang R, et al. Multi-Scale Mixed Attention Network for CT and MRI Image Fusion. *Entropy (Basel)* 2022;24:843.

45. Chattopadhyay A, Sarkar A, Howlader P, et al. Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 12-15 March 2018; Lake Tahoe, NV, USA. IEEE; 2018:839-47.
46. Hamza A, Attique Khan M, Wang SH, et al. COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization. *Front Public Health* 2022;10:1046296.
47. Desai S, Ramaswamy HG. Ablation-CAM: Visual Explanations for Deep Convolutional Network via Gradient-free Localization. *Proceedings of the IEEE/*
CVF Winter Conference on Applications of Computer Vision (WACV); 2020:983-91. Available online: https://openaccess.thecvf.com/content_WACV_2020/html/Desai_Ablation-CAM_Visual_Explanations_for_Deep_Convolutional_Network_via_Gradient-free_Localization_WACV_2020_paper.html
48. Abualigah L, Diabat A, Mirjalili S, et al. The Arithmetic Optimization Algorithm. *Comput Methods Appl Mech Eng* 2021;376:113609.
49. Abualigah L, Yousri D, Abd Elaziz M, et al. Aquila Optimizer: A novel meta-heuristic optimization algorithm. *Comput Ind Eng* 2021;157:107250.

doi: 10.21037/jmai-23-40

Cite this article as: Hirimutugoda YM, Silva TP, Wagarachchi NM. Handling the predictive uncertainty of convolutional neural network in medical image analysis: a review. *J Med Artif Intell* 2023;6:18.