A Review of Deep Neural Networkbased Uncertainty Quantification Methods for the Classification of Breast Cancer

Agha Salman Haider, Alighazi Siddiqui, Md Imran Alam,
Felipe de Castro Dantas Sales, Shams Tabrez Siddiqui,
V Vijayabhaskar, Ranju Lal and Harpreet Kaur

February 8, 2023

NeuroQuantology |September 2022 | Volume 20 | Issue 10 |Page 9702-9715| doi: 10.14704/nq.2022.20.10.NQ55947

A REVIEW OF DEEP NEURAL NETWORK-BASED UNCERTAINTY QUANTIFICATION METHODS FOR THE CLASSIFICATION OF BREAST CANCER

Agha Salman Haider[1], Alighazi Siddiqui[2], Md Imran Alam[3], Felipe de Castro Dantas Sales[4], Shams Tabrez Siddiqui[5], Vijayabhaskar V[6], Ranju Lal[7], Harpreet Kaur[8]
[1]Department of Information Technology and Security, College of Computer Science and Information Technology, Jazan University,Jazan, Saudi Arabia, aghasalmaanhaider@gmail.com
[2]Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia, ghazi.siddiqui@gmail.com
[3]Department of Computer and Network Engineering, College of Computer Science and Information Technology,Jazan University, Jazan, 45142, Saudi Arabia, imranmca04@gmail.com
[4]Institution/Campus: Must University / Florida, 1960 NE 5th Ave, Boca Raton, FL 33431, USA, felipecds@hotmail.com
[5]Department of Computer Science, College of Computer Science and Information Technology,Jazan University, Jazan, 45142, Saudi Arabia, stabrezsiddiqui@gmail.com
[6]Department of Science and HumanitiesR.M.K. Engineering CollegeKavaraipettaiThiruvallur DistrictTamil Nadu, vvbhaskar75@gmail.com
[7]Assistant Professor, Department of Psychology, Galgotias University, Greater Noida, UP, India, ranju.lal@galgotiasuniversity.edu.in
[8]Associate Professor, Lovely Professional University, Phagwara, 27633@lpu.co.in

ABSTRACT:

In recent years, deep learning-based technologies have become widely used in the medical area with remarkable success. The output of many of these methods, however, has excessive confidence levels and the majority of them cannot provide numerical guarantees. There is no way they could be effective, and they might even cause permanent harm. Therefore, the approximation of Bayesian and Ensemble learning techniques are considered as uncertainty quantification approaches to take on such a problem. In this study, we implement and assess three UQ models for categorising breast tumour tissues. A few examples of these techniques include the Bayesian Ensemble, the MCD Ensemble, and the Mont Carlo Dropout (MCD) approach. In addition, the present study takes into account a transfer learning technique and a pre-trained CNN in order to boost the classification's accuracy and remove the negative effects of the study's small data collection in Wisconsin Diagnostic Breast Cancer (WDBC). Novel performance criteria are used to assess estimated uncertainty, and the three proposed models are compared based on their capacity to quantify the reliability of classification. In the study, we conducted quantitative and qualitative analyses to indicate that models exhibit substantial ambiguity in misclassifications, which is critical for establishing the frequency of medical diagnosis hazards. Therefore, we hope to determine whether the deep neural network's output can be trusted by applying these new evaluation criteria. Further, the Bayesian Ensemble model's uncertainty quantification is shown to be more trustworthy through the analysis.

DOI Number: 10.14704/nq.2022.20.10.NQ55947 NeuroQuantology 2022; 20(10): 9702-9715

1. INTRODUCTION

Breast cancer is the leading female killer around the world. According to the World Health Organization, there would be 685,000 deaths from BC in 2020, with 2.3 million women being diagnosed. Timely and accurate identification of this type of cancer is crucial to the survival chances of patients; by the end of 2020, 7.8 million women will have survived after being diagnosed with BC during the preceding five years. Due to the fact that this cancer's diagnosis and treatment are receiving a lot of attention as of late, various initiatives have been launched to improve patient outcomes through earlier detection. Breast lumps are categorised as either carcinogenic (malignant) or noncancerous by the American Cancer Society (ACS) (benign). A prompt and accurate identification of the tumor's kind (benign or malignant) is crucial for determining the course of treatment from a medical perspective. Methods based on data mining and machine learning have helped healthcare practitioners by expediting the diagnostic process, improving accuracy, and providing a solid foundation for making treatment decisions as the amount of diagnostic datasets, including image and non-image data, has increased. As a result, many people worked to perfect and employ AI methods as reliable and efficient ancillary tools in breast cancer diagnosis, prognosis, and classification [1,2]. One of the primary steps in medical diagnosis is doing histological exams based on the properties of the removed tissue. Among the several techniques for detecting BC, tissue sampling ranks among the most crucial [3]. In traditional histological examinations, the type of tumour is diagnosed and the treatment method is determined based on the features derived from histopathological images, as well as the pathologist's prior knowledge and diagnosis. However, establishing the type of tumour process is laborious and its accuracy is unstable due to the wide variety of pictures available and the pathologist's experiences, workload, & working conditions [4][5]. Twenty-five percent or more of the time, pathologists disagree with each other [6]. As a result, it's vital to remember that complex diagnostic instances supported by medical data might play a distinguishing role in the diagnosis made by professionals. It is now possible to rapidly mine massive data for useful diagnostic information using AI detection, which can greatly speed up the screening process. It also has enormous potential in histopathological analysis, where it can reduce the number of incorrect diagnoses made due to human error and boost the effectiveness of a number of different diagnostic procedures. For instance, Nanglia et al. [7] used the CRISP-DM analytical methodology to predict BC using an Ensemble model based on three popular machine learning methods and achieved an accuracy of 78% as the best. At the same time, deep learning has been increasingly popular over the past decade for a wide range of applications, including medical analysis [8] and neuroscience-based emotion analysis based on brain signals [9,10]. Deep neural networks as well as transfer learning have been given substantial consideration due to their promising achievements in various diagnosis cases, including BC [11–14]. Prior research has mostly concentrated on classification models employing pre-trained DNNs and evaluated them using common performance measures like specific, sensitive, precision, accuracy, F-measure, and Matthews Coefficient Of correlation (MCC) [15]. This is due to the fact that the method of treating cancer depends on the type of tumour that has been identified. By considering these factors, it may become clear that a second medical opinion on the patient's condition is warranted and a referral made. The standard practise of seeking a second opinion can benefit from DNN's capacity to recognise diagnostic ambiguity. Determining diagnostic uncertainty in such a way can vastly enhance overall diagnostic performance [16]. The primary objective of this research is to apply DNN technology to the field of breast cancer diagnosis, which is one of the leading causes of death among humans, in order to not only classify tumour types, but also to determine the degree of classification certainty in order to lessen risks and enhance well-informed decision making. The following are examples of how this research adds to the existing body of knowledge:

- For the categorization of the WDBC (Breast Diagnostic Cancer Wisconsin) dataset, we will be comparing three deep learning-based models.
- Applying Bayesian methodology to generate an estimate of the model's output distribution.
- Utilizing Mont-Carlo dropout and transfer learning to enhance DNN estimation.
- Evaluating the suggested model's accuracy using the confusion matrix of doubtful conclusions.

The current study develops three models, which are all shown in Fig. 1. The methods used include Monte-Carlo Dropout, Ensemble, and Ensemble MC-dropout, as well as Bayesian theory for quantifying the ambiguity in DNN tumour type categorization and transfer learning for extracting key features. Additionally, performance criteria are used to provide an all-encompassing evaluation of uncertainty estimations in this study. As a result, both qualitative (through the use of the eyes) and quantitative (through the use of numbers) tests are performed (using uncertainty assessment criteria). Here are summaries of the paper's additional content. Deep uncertainty quantification is briefly reviewed in the context of its use in BC diagnostic investigations in Section 2. Methods using UQ are discussed in Section 3. Diagnostic uncertainty estimation quantitative assessment criteria are also outlined. Description and instructions for preparing the WDBC dataset take up the remainder of Section 3. Discussions and experimental data are presented in Section 4, and the study is wrapped up in Section 5.

2. RELATED STUDIES

Machine learning systems, given their role as a facilitator of superior decision-making, need to be capable of accurate and reliable reasoning under uncertain circumstances because of the prominence of ambiguity in estimating and making decisions. In spite of this, researchers have spent a lot of time focusing on deep learning applications with the goal of enhancing model performance and boosting prediction accuracy. To aid in the diagnosis of Alzheimer's disease, Salami et al. [17] suggested an Ensemble model based on a convolutional neural network, which showed impressive results using traditional performance metrics. In the meantime, other techniques have been developed to provide probabilistic estimations and enable DNNs to deal with uncertainty [18]. The accuracy of deep learning-based approaches and the predicted uncertainty were taken into account by Thagaard et al. [19], who looked at the impact of various distributional shifts in pathology datasets on these metrics. They tested three models for cancer detection in lymph nodes using MC-Dropout and Ensemble DNN, and they discovered that the Ensemble approach is a promising survey for prediction uncertainty. The mathematical tools for thinking rationally about model uncertainty are provided by Bayesian probability theory [20]. On the other hand, these methods are usually computationally intensive and difficult to implement [21,22]. Therefore, many researchers have proposed adopting a variety of approximate strategies to enhance the accuracy of uncertainty estimation by employing Bayesian neural networks [23,24]. Overfitting is a problem in DNNs, and so the dropout technique was developed to perform ambiguity in deep learning without increasing complexity [21, 25]. As a result, the dropout sampling technique has been utilised by numerous studies in order to quantify uncertainty. Gour and Jain [26] introduced the UA-ConvNet model, which uses a CNN for screening for COVID-19 illness from chest X-ray images and incorporates an approximation of ambiguity in the network's predictions. In this analysis, we use a customised version of the EfficientNet-B3 model in conjunction with the Monte Carlo (MC) dropout method. This research utilised three distinct evaluation metrics—standard deviation (SD), confidence interval (CI), and entropy—to compute the associated predictive uncertainty of the model. With Monte Carlo dropout and the neural network's association of input features, Fabi and Schneider [27] developed a technique to assess the link between the features and calculate an uncertainty score.Mojabi et al. [28] used deep learning to categorize tomographic microwave & ultrasound images and determine the type of concerning tissue in breast tumours. In order to generate tissue-like images and evaluate the uncertainty of the categorization of each pixel, the suggested method utilised a convolutional neural network (CNN) with a U-Net architecture. Finally, the author draws a parallel between her strategy and the Bayesian approaches she has previously explored, demonstrating substantial improvement in recovering tissue-type images. Khairnar et al. [29] employed a modified Bayesian-CNN model to categorise and assess the uncertainty of histological breast images. The suggested model featured a learnable parameter for each brain cell as part of an adaptive activation function. With this adaptive activation function, the loss function's performance could be dynamically adjusted, resulting in a faster and more accurate convergence than the Bayesian-CNN. Further, compared to CNN, the approach employed in this study significantly reduced the number of false-negative predictions (by around 38 percent). While there has been development in this area, a robust assessment of uncertainty is still a nascent field with room for investigation. Despite DNN's widespread application in the clinical and pathological diagnosis of a wide range of diseases like BC, the majority of research make recommendations based on ad hoc decisions and overlook the proposed model's dependability. Using a comparative methodology, Hassan Ibeni et al. [30] applied the Naive Bayes, Bayesian Networks, and Tree augmented Naive Bayes algorithms to three non-image data sets of BC and analysed the results. WEKA, a data mining tool, demonstrated that Bayesian Networks, with a precision of %97.281, were the most accurate classification technique. The CNN was utilised for semi-supervised learning by Sun et al. [31] to detect BC. The team came to the conclusion that DNNs, especially those used in medical diagnosis, required a massive amount of data in order to train properly. Faced with an insufficient dataset size, researchers have turned to transfer learning to either fine-tune the parameters of a previously trained network [32] or extract features from which to train a new classification model [33]. The application of transfer learning to automated medical diagnosis and direct inference of medical imaging data has been the subject of extensive research. Using convolutional neural network models that had been trained on non-medical images in ImageNet, Morid et al. [34] concluded that an approach that relied on a non-medical dataset may be useful in medical diagnosis. When it comes to creating accountable, transparent, and self-confident forecasting systems, Bayesian deep learning is indeed the leverage point for precise parameter uncertainty. Figure 2 is a high-level flowchart summarising the three models presented in this study to evaluate the ambiguity of breast tumour classification using Bayesian deep learning, transfer learning, as well as MC dropout.

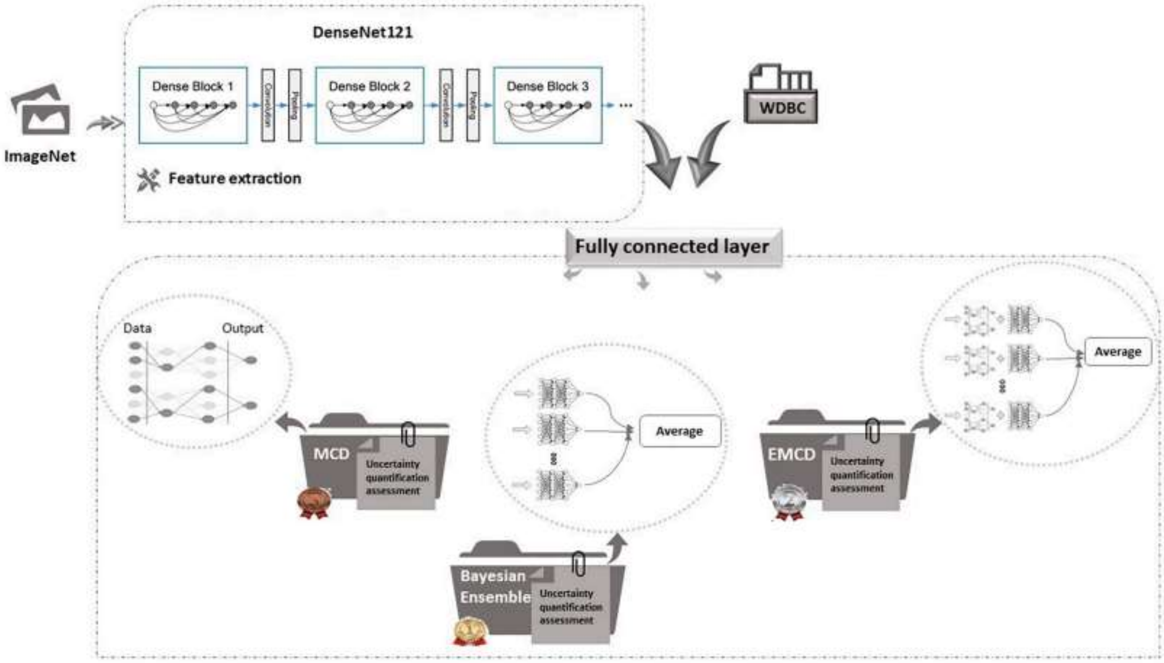


Fig. 1. An outline of the various approaches to quantifying uncertainty that have been proposed [52]

3. METHODOLOGY

Recently, there has been a lot of focus on extending DNNs to take into account probabilistic and uncertain approaches. UQ techniques, many of which are based on Bayesian approaches, have seen extensive use in the medical domain for the development of novel ways for efficient data processing. Prior distribution is set by the neural network parameters, and the probability of a parameters is computed using the training data to quantify the prediction's uncertainty.

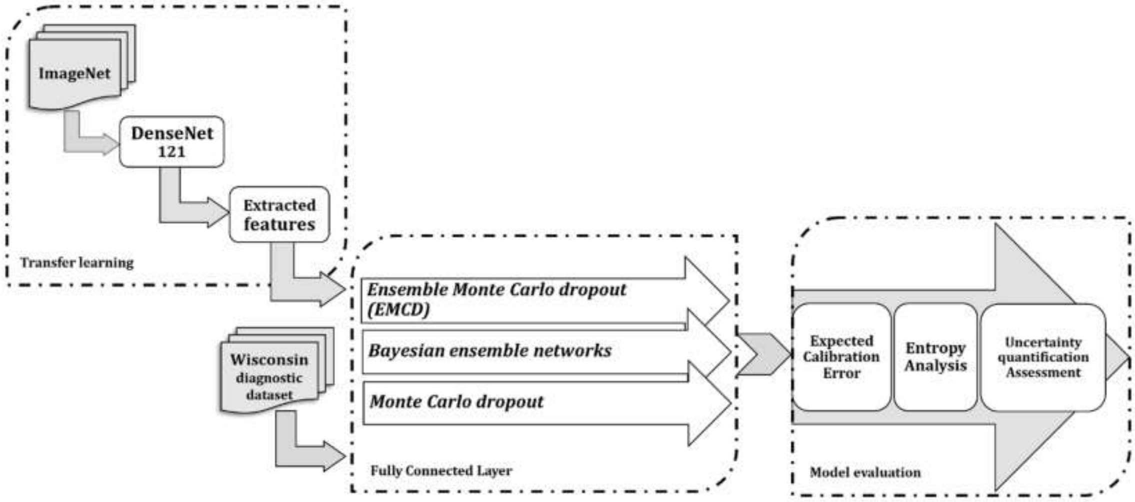


Fig. 2. Study procedure flowchart[52]

3.1. Uncertainty calculation techniques

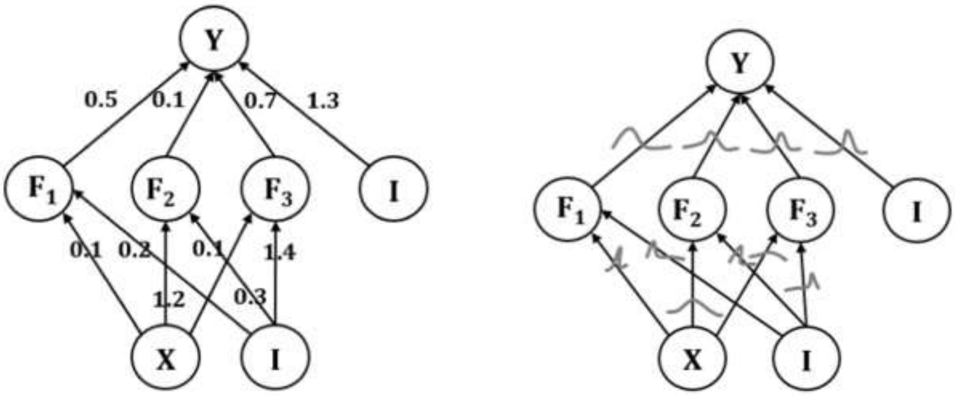
Through the use of specialised algorithms and statistical techniques, Bayesian neural networks simulate nonlinear and complicated issues, examine the current state, and make predictions about the future. Weights and biases are calculated points in traditional artificial neural networks where neurons in different layers communicate nonlinearly with one another (Fig. 3a). Parameters of neural networks are also provided as probability distributions in a Bayesian framework (Fig. 3b) and are trained with Bayesian inference. Bayes' theorem yields a formula for the posterior distribution, denoted by P(y|x,f), as follows: [20]:

$$P(Y|X,F) = \frac{P(F|X,Y)P(Y)}{P(F|X)}$$

The exact value of P (Y|X, F) in Eq. (1) cannot be calculated, but approximations can be made. Some sampling techniques are presented as a means to accurately estimate the posterior distribution. As a result, the entropy of the output is analysed as the model's uncertainty, and the mean and standard deviation of each approximation are used to describe them. In 1865, entropy was initially established in thermodynamics; Shannon [35] expanded its application to other sciences with a strong probabilistic bent. Entropy, as described by Shannon, is the degree to which outcomes are inherently unpredictable:

$$H = -\sum_{p_i} p_i \log_2 p_i$$

where pi is the probability distribution of i-th occurrence. Entropy measurement has been considered in various studies. Maasoumi and Racine [36] applied it to the field of finance to foretell time series and uncertainty. Integration of entropy and machine learning in the fields of medicine [37,38], neurology [39], as well as brain signal complexity analysis [40] has also received considerable interest. However, the Shannon entropy estimate of the probability density in Bayesian neural networks can be seen as a measure of prediction ambiguity, and therefore it was utilized in the present study.



(a) Deterministic (b) Probability distribution

Fig. 3. Network importance ratings[52]

3.1.1. Monte Carlo dropout (MCD)

Bayesian approximation is one of the methods used to quantify uncertainty. A major drawback of this approach is the time and effort required to compute the posterior distribution. The Bayesian approach, which makes use of Monte Carlo sampling, is proposed as an effective means of addressing this problem, since it allows for more precise measurement of uncertainty and thus more secure decision-making. To do so, Gal and Ghahramani devised the Bayesian dropout theory [21]. Bayesian theory-based machine learning techniques present the results as a probability distribution, whereas Mont Carlo dropout (MCD) minimises the computational load in estimations. In the testing step, dropout is used in conjunction with several stochastic forward passes (for T iterations) on the same input. Since the predictive mean as well as predictive entropy of this approximation probability density function of the output can be determined with an empirical estimator, this allows us to estimate the uncertainty of the input sample. Over a period of T iterations, the model's estimated mean is given by Eq (3):

$$\mu_{p_{red}} \approx \frac{1}{T}\sum_{t} p(y = c|x, \widehat{w_t})$$

x	Testing phase input
$\widehat{w_t}$	Config for the ith iteration of MC's iterative loop
$p(y = c\|x, \widehat{w_t})$	The possibility that y is a member of c

Equation (4) reflects the predicted entropy formed by the trained model and can be used as a quantifier of uncertainty in the model after training:

$$PE = -\sum_{c} \mu_{p_{red}} \log \mu_{p_{red}}$$

According to this definition of entropy, a larger level of PE indicates less confidence in the prediction's accuracy. Confidence and PE, in other words, are inversely connected.

3.1.2. Bayesian Ensemble networks

The method was developed with the goal of improving upon the performance accuracy of the individual techniques that make up the Ensemble. The goal of this approach is to accomplish a task by employing a group of Bayesian neural networks. To estimate the final distribution, we use the mean from Eq. (5) and the entropy from Eq. (6), where each network predicts a posterior probability independently. [41]:

$$\hat{p}(y|x) = \frac{1}{N} p_{\widehat{w_i}(y_i|x)}$$

$$PE_C = -\sum_{i\epsilon C} \hat{p}(y_i|x) \log \hat{p}(y_i|x)$$

$\widehat{w_i}$	The i-th network's set of parameters
C	C ∈ {classified, misclassified}∀i∈ {1,⋯,N}
N	Iteration count of neural networks in an Ensemble

If the predictions of both networks are roughly the same, as described by Eqs. (5) and (6), then the PE will be quite tiny.

3.2. Dataset

3.2.1. Dataset description

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset [47] is used to evaluate the accuracy of the suggested uncertainty calculation methods in Section 3.1. Researchers who employ machine learning strategies to evaluate the proposed approaches, including Ensemble classification [48], rely heavily on this data set as the gold standard for breast tumour classification. Fine Needle Aspiration (FNA) biopsy digital pictures were used to capture the features in this data collection. Histopathological images of breast tumour tissue can be analysed for certain characteristics of the cell nuclei, which are described in the WDBC. For each tissue sample, ten quantitative indices are retrieved from digital photographs of the tissue, with the intended values labelled as M (malignant) or B (benign) (benign). Of the 569 tumour tissue samples analysed, 357 (62.7%) were found to be benign, while the remaining 212 (37.3%) were malignant. In Table 1 you'll find feature names, descriptions, and typical update times. All data in this set is complete and accurate. Model overfitting, slow performance, and a high number of calculations are only some of the problems that plague the DNN approaches employed in this study. Additionally, the Bayesian DNN is sensitive to the quantity of the training data, with the uncertainty of the model decreasing as more data is used to train the model [49]. The WDBC dataset utilised in this investigation has 569 samples of tumour tissue, which raises concerns that the number of data points used to develop the models may be insufficient for correct network training and adequate parameterization, and therefore, the functionality of DNN. To alleviate these problems and compensate for the lack of a substantial dataset during DNN training, the medical diagnostics industry has turned to deep learning for feature extraction [50].

3.2.2. Dataset setup

In this research, we split the Wisconsin dataset in half, allocating 70% of the data for training and 30% for testing. Input is processed by a fully linked layer followed by a Soft-max operation. Using the RELU activation function and 300 epochs, we perform three quantitative strategies for measuring uncertainty. To acquire a rough distribution of test-phase prediction outcomes, we employ the Monte Carlo dropout sampling technique with a rate of 0.25 when the models were being trained. On top of that, the cross entropy cost function is optimised using ADAM, an optimising algorithm, and a learning rate of 0.001. In the MCD model, the neurons are distributed across three completely linked layers: layer 64, layer 256, and layer 512. Many tasks involve some degree of uncertainty that can be studied using ensemble models. And many other architectures exist for ensembles of deep networks in supervised learning. In this research, we use an ensemble of 30 networks, each of which was trained independently using the same dataset. This method's benefit lies in its ability to enhance learning power in parallel among ensemble members, leading to more robust distributions and outputs. Depending on the application, the proposed Ensemble networks can include anywhere from two to three layers. As an added bonus, the number of neurons in the fully connected layer is arbitrarily determined between (512) and (1024), (128) and (512), and (512) and (512). (128, 8). Let's imagine that one of the training Ensemble networks, when presented with some sample data, assigns a probability of x to the benign class and a probability of y to the malignant class. After the input has been classified, each of the remaining Ensemble networks uses its own classification to provide one among 30 possible results. The probability of the result is calculated. So, while the network isn't sure of its prediction, predictive entropy is large, and it's near zero whenever the outcome has been most likely assigned to a certain class. The final model, the EMCD model, is constructed in the same way as the Ensemble model, but the MC dropout purposive sampling is utilised to create each network.

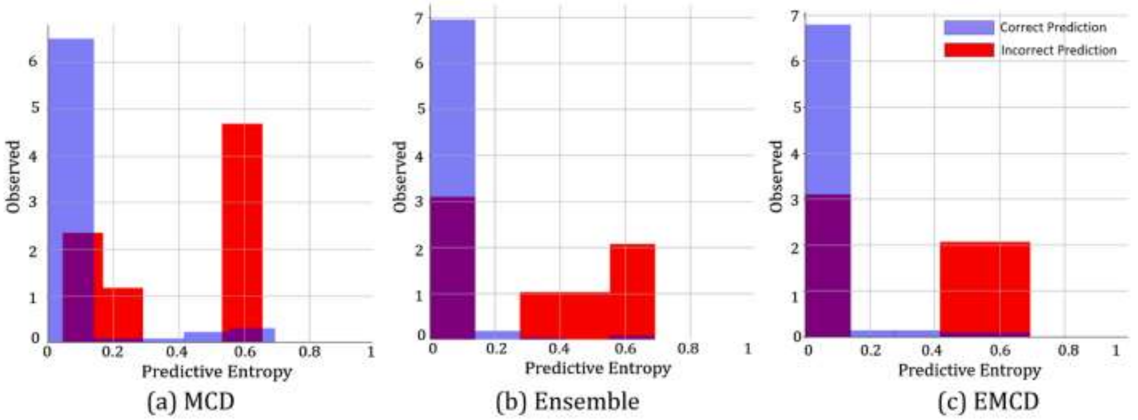
4. EXPERIMENTAL RESULTS

The boundary between the two classes determines whether the model's output falls into the certain or unsure class. The user must know how much entropy exposure is tolerable while seeking a medical diagnostic in order to successfully navigate parameters. Thresholds of entropy express the allowable range of error around a given prediction and provide a framework for defining this. For this reason, the entropy cutoff is not a discrete value but rather a continuous range from 0.1 to 0.9. The criteria for assessing the degree of prediction uncertainty were first presented in Section 3.4 and will be adjusted if the threshold moves within the allowed range. When the entropy threshold is raised, more confident predictions can be made since a larger margin of error is tolerated around the model's output. However, this comes at the expense of a smaller number of unreliable results.

4.1. Predictive entropy analysis

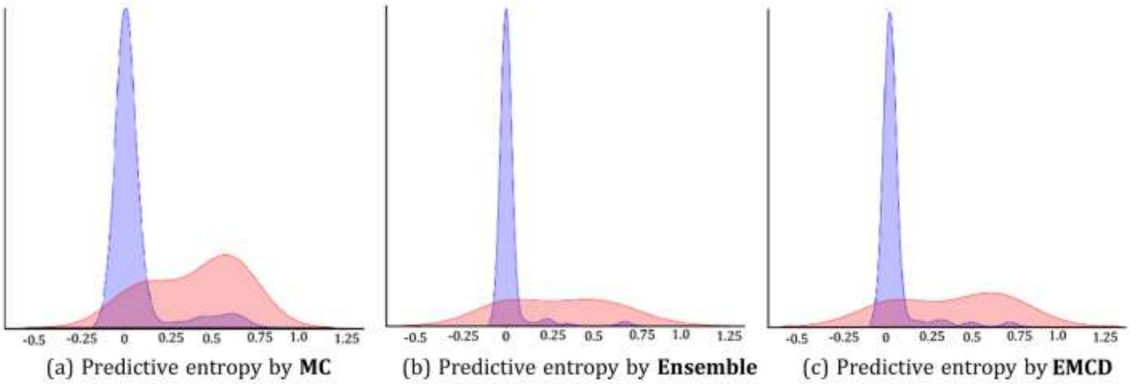
To compare the efficacy of three UQ approaches, we compute the entropy values and the accompanying predictive distributions. Taking into account the uncertainty factors supplied by the confusion matrix is important to us because we want to get the most out of the UQ approaches that have been discussed so far. The entropy of prediction during the testing phase of three examined UQ methods is shown in Fig. 4 as histogram diagrams. As a result, TC samples, which have been shown to be accurate, cluster to the left (blue part). Alternatively, misclassified samples (TU) tend to cluster to the right of the horizontal axis (red part).

Indicative of the accuracy of these models is a difference in entropy between the TC and TU, two crucial elements of the uncertainty confusion matrix. There are some false and certain predictions (FC) and true and uncertain predictions (FU) among the TC and TU, and they were labelled as the result of the model's inability to evaluate predictions confidence. Figure 5 depicts the posterior distribution of the models' predictions. Using Eq (4). The true categorised samples are represented by the blue graph, whereas the incorrectly classified samples are shown in the red graph. When the output has a wider distribution, it means the model is less reliable. Higher estimated uncertainty regarding the probable tumour type causes a larger range of predicted outcomes. If the model's entropy is very high around the estimated mean, it is best to consult a second expert for confirmation of the diagnosis.



(a) MCD (b) Ensemble (c) EMCD

Fig. 4 Predictive entropy histogram plots for three distinct types of UQ are displayed[52]



(a) Predictive entropy by **MC** (b) Predictive entropy by **Ensemble** (c) Predictive entropy by **EMCD**

Fig. 5 Variation in predicted model results due to aforementioned sources of uncertainty[52]

5. CONCLUSION

Using the numerical data from the Wisconsin dataset, the authors of this study assessed the efficacy of many quantitative methodologies with high uncertainty for determining the kind of breast tumour. By combining transfer learning with the Bayesian Dropout method, DNNs can increase their prediction and binary classification abilities. Results from using three different uncertainty methods—MC dropout, Bayesian Ensemble, and Ensemble MC-dropout—are described and contrasted. To evaluate the uncertainty, estimate in prediction and select the optimal UQ approach, a novel confusion matrix incorporating many performance criteria is utilised. The findings of this thorough evaluation show that deep Ensembles are more reliable in terms of performance and calibration, and they also help us detect inaccurate predictions more effectively. Important to keep in mind is that Ensemble models are always considered a sampling method to improve the model's accuracy. For this reason, the diagnostic options they offer are more solid. Our research shows that this network type has the most promising quantification uncertainty performance. With an accuracy of 0.981, the ensemble of Bayesian neural networks used in this study proves that Bayesian theory may be successfully implemented in neural network-based models employing ensemble methodologies for quantifying degrees of certainty. An efficient method for estimating the uncertainty of models is to use Bayesian theory and related approximations in the context of prediction models to obtain the distribution of the output. This survey can be used to quantify and easily evaluate the credibility of the outputs in large, ambiguous areas. However, solutions that make these methods more accessible, like the Dropout approximation (as evidenced by the findings of the provided method), may be the cause of worse model reliability and additional uncertainty. As a result, this may reduce the reliability of the forecast. Future research could look into how the effect of the unbalanced dataset encountered in this study has on the process of quantifying the uncertainty of predictions, or the uncertainty produced from algorithms like transfer learning approaches and MC-dropout.

REFERENCES

1 S. Bharati, P. Podder, M. Hossain Mondal, Artificial neural network based breast cancer screening: a comprehensive review, Int. J. Computer Information Syst. Ind. Manage. Applications 12 (2020) 125–137.

2 M.H. Alshayeji, H. Ellethy, R. Gupta, Computer-aided detection of breast cancer on the Wisconsin dataset: an artificial neural networks approach, Biomed. Signal Process. Control 71 (2022), 103141.

3 https://www.nationalbreastcancer.org/breast-cancer-biopsy.

4 K.H. Allison, L.M. Reisch, P.A. Carney, D.L. Weaver, S.J. Schnitt, F.P. O'Malley, B. M. Geller, J.G. Elmore, Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel, Histopathology 65 (2) (2014) 240–251, https://doi.org/10.1111/his.12387.

5 X. Zhou, C. Li, M.M. Rahaman, Y. Yao, S. Ai, C. Sun, G. Wang, Y. Zhang, M. Li, X. Li, T. Jiang, D. Xue, S. Qi, Y. Teng, A comprehensive review for breast histopathology image analysis using classical and deep neural networks, IEEE Access 8 (2020) 90931–90956, https://doi.org/10.1109/ACCESS.2020.2993788.

6 J.G. Elmore, G.M. Longton, P.A. Carney, B.M. Geller, T. Onega, A.N. Tosteson, H. D. Nelson, M.S. Pepe, K.H. Allison, S.J. Schnitt, Diagnostic concordance among pathologists interpreting breast biopsy specimens, J. Am. Med. Assoc. 313 (11) (2015) 1122–1132, https://doi.org/10.1001/jama.2015.1405.

7 S. Nanglia, M. Ahmad, F.A. Khan, N.Z. Jhanjhi, An enhanced Predictive heterogeneous ensemble model for breast cancer prediction, Biomed. Sig. Process. Control 72 (A) (2022), https://doi.org/10.1016/j.bspc.2021.103279.

8 S. Shamshirband, M. Fathi, A. Dehzangi, A.T. Chronopoulos, H. Alinejad-Rokny, A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues, J. Biomed. Inform. 113 (2021), 103627, https://doi.org/10.1016/j.jbi.2020.103627.

9 S. Bagherzadeh, K. Maghooli, A. Shalbaf, A. Maghsoudi, Recognition of emotional states using frequency effective connectivity maps through transfer learning approach from electroencephalogram signals, Biomed. Signal Process. Control 75 (2022), 103544, https://doi.org/10.1016/j.bspc.2022.103544.

10 S. Aydın, Deep learning classification of neuro-emotional phase domain complexity levels induced by affective video film clips, IEEE J. Biomed. Health

Informatics 24 (6) (2019) 1695–1702, https://doi.org/10.1109/JBHI.2019.2959843.

11 S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, R. Deters, Classification of histopathological biopsy images using ensemble of deep learning networks, arXiv preprint arXiv:1909.11870, 2019, 10.1145/3306307.3328180.

12 R. Mehra, Breast cancer histology images classification: training from scratch or transfer learning? Information Commun. Technol. Express 4 (2018) 247–254, https://doi.org/10.1016/j.icte.2018.10.007.

13 M. Jannesari, M. Habibzadeh, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi, I. Hajirasouliha, Breast Cancer Histopathological Image Classification: A Deep Learning Approach, in: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2018, pp. 2405–2412, https://doi.org/ 10.1109/BIBM.2018.8621307.

14 W. Zhi, H. Yueng, Z. Chen, S.M. Zandavi, Z. Lu, Y.Y. Chung, Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images, in: Proceedings of the 24th International Conference on Neural Information Processing, 2017, pp. 669–676, https://doi.org/10.1007/978-3-319-70093-9_71.

15 C. Kaushal, S. Bhat, D. Koundal, A. Singla, Recent trends in computer assisted diagnosis (CAD) system for breast cancer diagnosis using histopathological image, Innovation Res. BioMed. Eng. 40 (2019) 211–227, https://doi.org/10.1016/j. irbm.2019.06.001.

16 M. Raghu, K. Blumer, R. Sayres, Z. Obermeyer, B. Kleinberg, S. Mullainathan, J. Kleinberg, Direct uncertainty prediction for medical second opinions, in: Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, 97(2019) 5281–5290.

17 F. Salami, A. Bozorgi-Amiri, G.M. Hassan, R. Tavakkoli-Moghaddam, A. Datta, Designing a clinical decision support system for Alzheimer’s diagnosis on OASIS-3 data set, Biomed. Sig. Process. Control 74 (2022), 103527, https://doi.org/ 10.1016/j.bspc.2022.103527.

18 M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. Rajendra Acharya, V. Makarenkov, S. Nahavandi, A review of uncertainty quantification in deep learning: Techniques, applications and challenges, Information Fusion 76 (2021) 243–297, https://doi. org/10.1016/j.inffus.2021.05.008.

19 J. Thagaard, S. Hauberg, B. van der Vegt, T. Ebstrup, J. D. Hansen, A. B. Dahl, Can You Trust Predictive Uncertainty Under Real Dataset Shifts in Digital Pathology ?, Springer International Publishing, Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention, Lima, Peru, 12261(2020) 824-833. Doi: 10.1007/978-3-030-59710-8_80.

20 J.M. Bernardo, A.F. Smith, Bayesian theory, John Wiley & Sons 405 (2009).

21 Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: International Conference on Machine Learning, The Proceedings of Machine Learning Research (2016, June) 1050-1059.

22 B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, in: Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA (2017).

23 J. M. Herńandez-Lobato, R. P. Adams. Probabilistic backpropagation for scalable learning of Bayesian neural networks, in: Proceedings of the 32nd International Conference on Machine Learning, Lille, France, The Proceedings

of Machine Learning Research 37 (2015) 1861-1869.

24 J. T. Springenberg, A. Klein, S. Falkner, F. Hutter, Bayesian optimization with robust Bayesian neural networks, in: Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain (2016).

25 N. Srivastava, G. Hinton, Dropout: a simple way to prevent neural networks from overftting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

26 M. Gour, S. Jain, Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification, Comput. Biol. Med. 140 (2022), 105047, https://doi.org/ 10.1016/j.compbiomed.2021.105047.

27 K.Fabi, J. Schneider, on feature relevance uncertainty: a Monte Carlo Dropout sampling approach, arxiv: 2008.01468v1 [cs.lg] 4 Aug (2020).

28 P. Mojabi, V. Khoshdel, J. Lovetri, Tissue-Type classification with uncertainty quantification of microwave and ultrasound breast imaging: a deep learning approach, IEEE Access 8 (2020) 182092–182104, https://doi.org/10.1109/ ACCESS.2020.3027805.

29 P. Khairnara, P. Thiagarajana, S. Ghoshb, A modified Bayesian convolutional neural network for breast histopathology image classification and uncertainty quantification, arXiv:2010.12575v1 [cs.CV] 7 Oct. 2020.

30 W.N.L.W. Hassan Ibeni, M.Z.M. Salikon, A. Mustapha, S.A. Daud, M.N.M. Salleh, Comparative analysis on Bayesian classification for breast cancer problem, Bull. Electrical Eng. Informatics 8 (4) (2019) 1303–1311, https://doi.org/10.11591/ EEI.V8I4.1628.

31 W. Sun, T.-L.B. Tseng, J. Zhang, W. Qian, Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data, Computerized Med. Imaging Graphics 57 (2017) 4–9, https://doi.org/10.1016/j. compmedimag.2016.07.004.

32 H. Ahmad, S. Ghuffar, K. Khurshid, Classification of breast cancer histology images using transfer learning, in: Proceedings of the 16th International Bhurban Conference on Applied Sciences and Technology, Islamabad, Pakistan (2019), Doi: 10.1109/IBCAST.2019.8667221.

33 S. Saxena, S. Shukla, M. Gyanchandani, Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology, Int. J. Imaging Syst. Technol. (2020) 1–15, https://doi.org/10.1002/ima.22399.

34 M.A. Morid, A. Borjali, G. Del Fiol, A scoping review of transfer learning research on medical image analysis using ImageNet, Comput. Biol. Med. 128 (2021), 104115, https://doi.org/10.1016/j.compbiomed.2020.104115.

35 C.E. Shannon, A mathematical theory of communication, ACM SIGMOBILE mobile computing and communications review 5 (1) (2001) 3–55, https://doi.org/ 10.1145/584091.584093.

36 E. Maasoumi, J. Racine, Entropy and predictability of stock market returns, J. Econometrics 107 (1–2) (2002) 291–312, https://doi.org/10.1016/S0304-4076 (01)00125-7.

37 H. Lou, H. Liu, Z. Chen, B. Dong, J. Xu, Bio-process inspired characterization of pregnancy evolution using entropy and its application in preterm birth detection, Biomed. Signal Process. Control 75 (2022), 103587, https://doi.org/10.1016/j. bspc.2022.103587.

38 I. Boukhennoufa, X. Zhai, V. Utti, J. Jackson, K.D. McDonald-Maier, Wearable sensors and machine learning in post-stroke rehabilitation assessment: a systematic review, Biomed. Signal Process. Control 71 (2022), 103197,

https://doi.org/ 10.1016/j.bspc.2021.103197.

39 S. Aydın, Ç. Güdücü, F. Kutluk, A. Oniz, ¨ M. Ozg ¨ oren, ¨ The impact of musical experience on neural sound encoding performance, Neurosci. Lett. 694 (2019) 124–128, https://doi.org/10.1016/j.neulet.2018.11.034.

40 F.H. Çetin, M. Barıs¸ Usta, S. Aydın, A.S. Güven, A Case Study on EEG Analysis: embedding entropy estimations indicate the decreased neuro-cortical complexity levels mediated by methylphenidate treatment in children with ADHD, Clin. EEG Neurosci. (2021), https://doi.org/10.1177/15500594211064008, 15500594211064008.

41 J. Van Amersfoort, L. Smith, Y.W. Teh, Y. Gal, Uncertainty estimation using a single deep deterministic neural network, in: International conference on machine learning, 2020, pp. 9690–9700.

42 J. Wu, X.Y. Chen, H. Zhang, L.D. Xiong, H. Lei, S.H. Deng, Hyperparameter optimization for machine learning models based on Bayesian optimization, J. Electron. Sci. Technol. 17 (1) (2019) 26–40, https://doi.org/10.11989/ JEST.1674-862X.80904120.

43 V. Cheplygina, M. de Bruijne, J.P. Pluim, Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis, Med. Image Anal. 54 (2019) 280–296, https://doi.org/10.1016/j.media.2019.03.009.

44 S. Sharma, R. Mehra, Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images - A comparative insight, J. Digit. Imaging 33 (3) (2020) 632–654, https://doi.org/10.1007/s10278- 019-00307-y.

45 G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu (July 2017) 4700–4708, Doi: 10.1109/ TPAMI.2019.2918284.

46 H. Asgharnezhad, A. Shamsi, R. Alizadehsani, A. Khosravi, S. Nahavandi, Z.A. Sani, D. Srinivasan, S.M.S. Islam, Objective evaluation of deep uncertainty predictions for covid-19 detection, Sci. Rep. 12 (1) (2022) 1–11, https://doi.org/10.1038/ s41598-022-05052-x.

47 https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29.

48 M. Hosni, I. Abnane, A. Idri, J.M. Carrillo de Gea, J.L. Fern´ andez Alem´ an, Reviewing ensemble classification methods in breast cancer, Computer Methods Programs Biomed. 177 (2019) 89–112, https://doi.org/10.1016/j. cmpb.2019.05.019.

49 R. Zheng, S. Zhang, L. Liu, Y. Luo, M. Sun, Uncertainty in Bayesian deep label distribution learning, Appl. Soft Comput. 101 (2021), 107046, https://doi.org/ 10.1016/j.asoc.2020.107046.

50 H. Zerouaoui, A. Idri, Deep hybrid architectures for binary classification of medical breast cancer images, Biomed. Signal Process. Control 71 (2022), 103226, https:// doi.org/10.1016/j.bspc.2021.103226.

51 C. Guo, G. Pleiss, Y. Sun, and K.Q. Weinberger, On calibration of modern neural networks, in: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 70 (2017) 1321-1330. arXiv:1706.04599v2 [cs.LG] 3 Aug 2017.

52 Fatemeh Hamedani-KarAzmoudehFar, Reza Tavakkoli-Moghaddam, Amir Reza Tajally, Seyed Sina Aria,Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods,Biomedical Signal Processing and Control,Volume 79, Part 1,2023,104057,ISSN 1746-

8094,https://doi.org/10.1016/j.bspc.2022.104057.