Indonesian Journal of Electrical Engineering and Computer Science
Vol. 30, No. 2, May 2023, pp. 1167~1177
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v30.i2.pp1167-1177

Bayesian deep learning methods applied to diabetic retinopathy disease: a review

Halbast Rashid Ismail[1], Masoud Muhammed Hassan[2]
[1]Technical College of Informatics-Akre, Duhok Polytechnic University, Duhok, Iraq
[2]Department of Computer Science, Faculty of Science, University of Zakho, Duhok, Iraq

Article Info

Article history:

Received Sep 7, 2022
Revised Jan 17, 2023
Accepted Jan 19, 2023

Keywords:

Bayesian deep learning
Diabetic retinopathy
Markov chain monte carlo
MC dropout
Variational inference

ABSTRACT

Diabetic retinopathy (DR) is a complication of diabetes that cause retinal damage; therefore, it is a leading cause of blindness. However, early detection of this disease can dramatically reduce the risk of vision loss. The main problem of early DR detection is that the manual diagnosis by ophthalmology is time-consuming, expensive, and prone to misdiagnosis. Deep learning (DL) models have aided in the early diagnosis of DR, and DL is now frequently utilized in DR detection and classification. The main issues with classical DL models is that they are incapable to quantify the uncertainty in the models, thus they are prone to make wrong decisions in complex cases. However, Bayesian deep learning (BDL) models have recently evolved as unified probabilistic framework to integrate DL and Bayesian models to provides an accurate framework to identify all sources of uncertainty in the model. This paper introduces BDL and most recent research that used BDL approaches to treat diabetic retinopathy are reviewed and discussed. A thorough comparison of the existing Bayesian approaches in this topic is also presented. In addition, available datasets for the fundus retina, which is often employed in DR, are provided and reviewed.

Corresponding Author:

Halbast Rashid Ismail
Technical College of Informatics-Akre, Duhok Polytechnic University
Duhok, Iraq
E-mail: halbast.rashid@dpu.edu.krd

1. INTRODUCTION

Deep learning (DL) models have attained excellent results in a diversity of problems, in classification large-scale images [1], processing of natural language [2], and segmentation of the medical images [3]. However, standard approaches are discovered to produce more predictions than necessary, which means they are not calibrated correctly [4]. In tasks of classification, for example, a badly calibrated network can put a high possibility block over one of classes, even if the expected class is wrong, whilst a classification that well-calibrated will put a lower possibility block on the unsure classes. Thus, to mitigate these risks several methods have been suggested [5]. Among them is the Bayesian model, which offers a rigorous framework for training and assessing neural networks that are aware of uncertainty, as well as supporting the development of the learning algorithms in general.

Bayesian deep learning (BDL) provides a pragmatic method to merging modern deep learning models with Bayesian probability theory. Instead of providing point estimates for the network weights, BDL offers the entire distribution and may quantify their uncertainty. Figure 1 depicts the correspondence between deep learning concepts for point estimation neural networks and their counterparts in Bayesian neural networks (BNNs). BDL is interested in the development of tools and techniques for quantifying uncertainty when the model is uncertain, and provides a probabilistic inference for the problem under study. BDL has already proved its success to play an essential role in applications for example diagnostics medical [6], computer vision [7], applied sciences [8], and autonomous driving [9].

Deep neural networks have also gained popularity in recent years for automatic classification of diabetic rretinopathy (DR) [10]. DR is a complication of diabetes that cause retinal damage, and it as non-proliferative DR (NPDR) or pro liferative DR (PDR) based on the presence or absence of new abnormal vessels. NPDR can also be classified as severe, moderate and mild. These levels can be identified based on the risk of their development [11]. Figure 2 and Table 1 illustrate the clinical international DR scale [12], where Figures 2(a)-(e) illustrate different stages of DR. Recently, the focus of interest has shifted for the classification task to developing robust deep learning models, most typically using BDL approaches that approximates the posterior distribution of a BNNs in a computationally scalable way. Previous studies have looked at a diversity of topics from studying the advantages of model uncertainty estimations [13] to developing computational robust methods [14]. Despite the diversity of various studied algorithms [13], the datasets used were benchmark datasets. The question of whether these algorithms generalize to clinical datasets remains unanswered. Furthermore, current research has mainly concentrated on the classification of DR using binary classifier approaches, namely ‘healthy versus any DR' or ‘referable versus non-referable (RDR)'. However, there has been a change toward the five-class suggested international DR classifier system (PIRC) in clinically oriented methods [15], as given in Table 1.

The importance of using BDL arises when dealing with image recognition in nonlinear dynamical systems [16]. For example, when dealing with medical complex images, the perception from raw images will be transformed to multiple layers using simple nonlinear methods. Yet to control the uncertainty in the data and the model itself we need a more effective approach, and only BDL can deal with these two hard tasks simultaneously. Therefore, BDL models are able to provide a probabilistic framework to manage and quantify different uncertainties in the model.

The remainder of this paper is organized as follows: section 2 provides background of classical and Bayesian deep learning methods. The literature review is presented in section 3, while section 4 provides discussion and comprehensive comparison on the reviewed papers. Finally, the main conclusions are drawn in section 5.

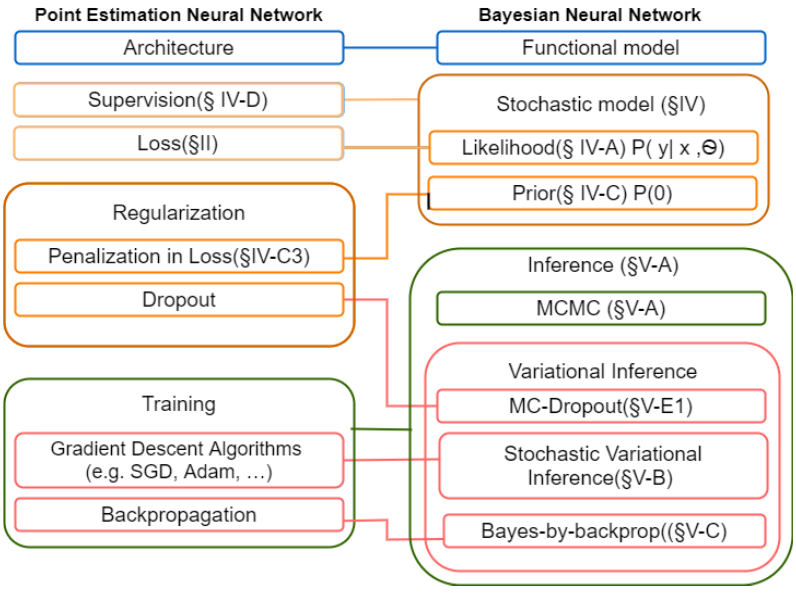
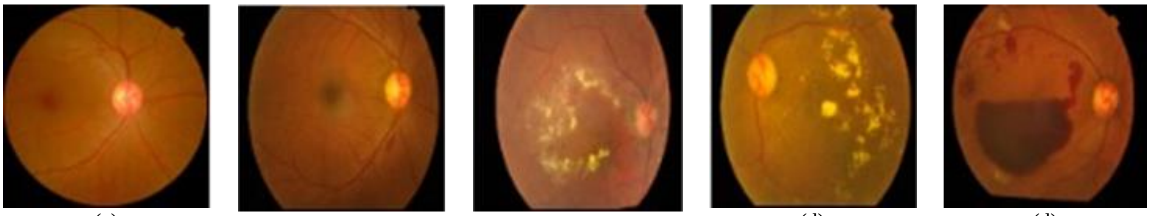


Figure 1. An illustration of the matching between concepts used in DL for neural networks with point estimation and their counterparts in BNNs [17]

Table 1. Diabetic retinopathy severity scale. Non-referable DR (R_0 and R_1) and referable DR (R_2, R_3, R_4)

Grade	Description
(R0) – ‘No DR’	Indicates that there is no DR.
(R_1) – ‘Mild NPDR’	Only microaneurysms are present.
(R_2) – ‘Moderate NPDR’	Is more than just microaneurysms, but it is not as bad as severe NPDR.
(R_3) – ‘Severe NPDR’	One or more of the following; >20 intraretinal hemorrhages Beading of the veins, Microvascular anomalies in the intraretinal space. There is no evidence of PDR.
(R_4) – ‘Proliferative DR’	One or both of the following options: Neovascularization, Pre-retinal/vitreous.



(a) (b) (c) (d) (d)

Figure 2. Different stages of DR: (a) no DR, (b) mild, (c) moderate, (d) severe, and (e) PDR

2. BACKGROUND
2.1. Datasets
There are several publicly available Retinal image datasets that can be obtained. These datasets are widely used by researchers to develop DL models and to use for comparing different diabetic retinopathy classification techniques. The common available datasets are EyePACS [18], DDR [19], DIARETDB [20], STARE [21], Messidor [22], RFMiD [23], and APTOS [24], HEIMED [25], e-ophtha [26], ROC [27] and DRIVE [28]. In most recent studies some private datasets are also leveraged to improve the accuracy of pre trained models [29]. Table 2 illustrates an overview of all open-source datasets of DR [30].

Table 2. Details of diabetic retinopathy datasets

Dataset	Image Count	Image Size (px)
EyePACS from kaggle 2015 [19]	88,702	(433×289) to (5,185×3,456)
DDR [20]	13,673	Varies
ODIR [31]	10,000	-
APTOS 2019 [25]	3,660	Varies
RFMiD [24]	3,200	(2,144×1,424)
Messidor [23]	1,200	(1,440×960) to (2,304×1,536) 24-bit
Messidor-2 [23]	1,784	(1,440×960) to (2,304×1,536) 24-bit
IDRiD [32]	516	(4,288×2,848)
DIARETDB0 [21]	130	(1,500×1,152) 24-bit
DRIVE [29]	40	(565×584) 24-bit
DIARETDB1 [33]	89	(1,500×152) 24-bit

2.2. Deep learning
Deep learning is a branch of machine learning based on artificial neural networks with multiple hidden layers [34]. It can be used as supervised or unsupervised learning, and it provides results more accurate than traditional machine learning algorithms. The rapid development of DL methods in different fields of science showed its successful due to advances in available hardware and software algorithms and the large data sets availability [35]. DL can learn to extract features of data during training and then incorporate this knowledge into the neural network's parameters in the form of weights and biases. DL models come with various types of architectures [35], including recurrent neural networks (RNN), convolution neural networks (CNN), deep belief networks (DBN), long short-term memory (LSTM), deep stacking networks (DSN). Here we mainly focus on the CNN model as it plays a significant role in the field of medical imaging [36]. A CNN is a type of image processing architecture in which different layers execute different functions as needed. The convolution layer is used to create pattern identification filters, while the Relu layer is used to normalize image values to be in a positive vector. The flatten and fully connected layers are the next layers, where the flatten turns the 2d array into a 1d array, which is thereafter transferred to the fully connected (FC) layers, in which it works similarly as an artificial neural network. Many different CNN architectures have been developed over the years, some of most common architectures exist LaNet, EffcientNet, VGGNet, ResNet and others [37]. CNNs are utilized in various ML applications, such as computer vision, video processing, time series prediction, and natural language processing (NLP), in addition to being the most prevalent approach to images processing [38]. Recently, CNN has achieved popularity in Bayesian DL, and it has been used in various applications, including medical imaging [39], text identification [40], and studies of genome [41]. Figure 3 demonstrates the CNNs model's basic architecture.

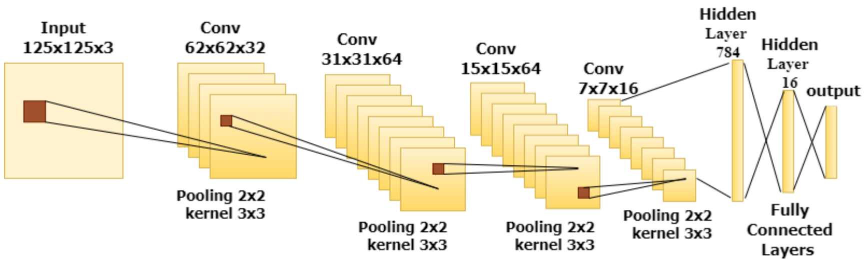


Figure 3. Basic architecture of the CNNs model

2.3. Bayesian deep learning

BDL denotes to probabilistic DL, which primarily relies on the Bayes theorem. In Bayesian techniques, the likelihood of the data and prior "expert" knowledge are used to create posterior distributions, which can indicate various levels of modal uncertainty [42]. It is noteworthy that there is another phrase used in the literature is Bayesian neural networks (BNN). Bayesian processing for neural networks (NN) is a public subject of utilizing Bayesian techniques in NN models where by Bayesian inference is utilized for fine-tuning activation functions and weights by using Bayesian optimization inference. On the other hand, to address various types of model uncertainties, BDL is concerned with determining the posterior distribution of weights. BDL has several benefits over classical DL [43]. Any shape of probabilistic distributions like Gaussian, beta, gamma or exponential can take the likelihood and prior distributions. Figure 4 illustrates the difference between classical and Bayesian NNs.

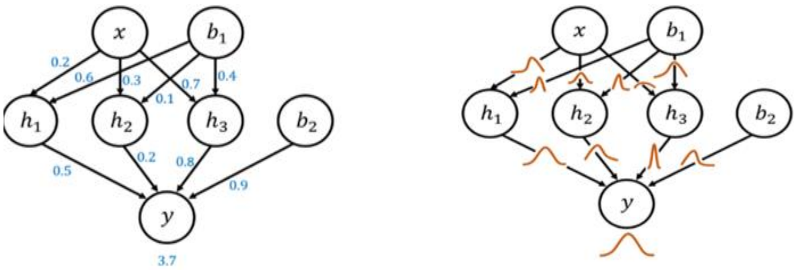


Figure 4. Standard neural network (left) and Bayesian neural network (right) [42]

There are two types of uncertainty that BDL can handle in general: Aleatoric uncertainty and epistemic uncertainty [44] see Figure 5. Aleatoric uncertainty is the type of complicated uncertainty in data "known as data uncertainty" that causes uncertainty in predictions [45]. This kind of uncertainty isn't a characteristic of the model, or rather is ingrained property of the data distribution; hence, it is irreducible. On the other hand, Epistemic uncertainty (as well-known as knowledge uncertainty) results from insufficient knowledge in the model framework. Epistemic uncertainty is manageable, and the Bayesian approach is an effective means of handling this kind of uncertainty without overfitting [42]. Seeking to quantify epistemic uncertainty, one can select different models to respond various questions in model-based prediction.

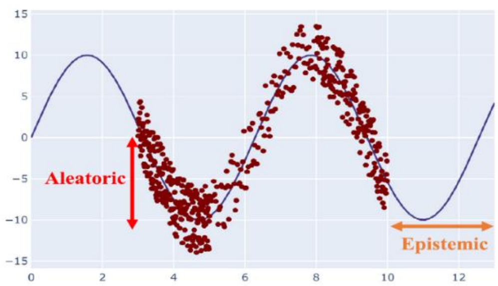


Figure 5. A graph shows the key distinctions between epistemic and aleatoric uncertainty [45]

2.3.1. Bayesian modelling

Given the inputs of training $\mathrm{x} = \{x_1, \dots, x_n\}$ and their associated outputs $y = \{y_1, \dots, y_n\}$ in regression of Bayesian (parametric), one can find the parameters w in the function$y = f_w(\mathrm{x})$ hat are most likely to have produced our outcomes can be identified. Which parameters are most probably to have created our data? Using Bayesian method, we will place some prior distributions over parameters space, p(w). As a result of our prior belief, this distribution describes which parameters might have created our data before we notice any points of data [46]. This distribution will be changed when some data is observed to represent the most and less probable parameters based on the observed data points. In order this, also need to determine a distribution of likelihood $p(y|\mathrm{x}, w)$ A probability model that generates the outputs from the inputs given some parameter settings w. For multi-classification tasks, we may suppose a function of softmax [47], as follows:

$$p(y = d \mid \mathrm{x}, w) = \frac{\exp(\mathrm{f}_w(\mathrm{x}))}{\sum_d \exp(\mathrm{f}_w(\mathrm{x}))} \tag{1}$$

or Gaussian likelihood for regression as shown in (2):

$$p(\mathrm{y} \mid \mathrm{x}, w) = \mathcal{N}(\mathrm{y}; \mathrm{f}_w(\mathrm{x}), \tau^{-1} I) \tag{2}$$

with the precision of modelτ. This might be considered a model output corruption due to variance and observation noiseτ^{-1}.

In (3), We then use Bayes' theorem to find the posterior distribution across the space of parameters given data *(X, Y)*:

$$p(w \mid \mathrm{X}, \mathrm{Y}) = \frac{p(\mathrm{Y}|\mathrm{X}, w) p(w)}{p(\mathrm{Y}|\mathrm{X})} \tag{3}$$

given our observed data, this distribution depicts the most potential function parameters. By integrating, we may use it to predict the outcome for a new input point x^* as shown in (4),

$$p(\mathrm{y}^* \mid \mathrm{x}^*, \mathrm{X}, \mathrm{Y}) = \int p(\mathrm{y}^* \mid \mathrm{x}^*, w) p(w \mid \mathrm{X}, \mathrm{Y}) \mathrm{d}w \tag{4}$$

we refer to this process as inference.

In (5), the normalizer, also known as model evidence, is a crucial element in posterior evaluation:

$$p(\mathrm{Y} \mid \mathrm{X}) = \int p(\mathrm{Y} \mid \mathrm{X}, w) p(w) \mathrm{d}w \tag{5}$$

due to the difficulty of computing the evidence (integrals), calculating the posterior of Bayesian and sampling from it are typically obstinate. Instead, one of the following Bayesian inference techniques are generally used [45] to integrate the above posterior distribution:

A. Markov chain Monte Carlo

Markov chain Monte Carlo (MCMC) [48] is an effective sampling methods in Bayesian inference. It starts by taking a randomly drawn value Z_t from the distribution $p(Z_t|\mathbf{x})$. Then, a random transition applies as (6):

$$Z_t \sim \mathrm{g}(Z_t \mid Z_{t-1}, \mathbf{x}) \tag{6}$$

this transition factor is selected and reiterated T times and the result is random variable that converges in the distribution to the precise posterior. However, in DL models, the big volume of data and sheer size of model parameters make the use of MCMC techniques too slow to be used [49]. The number of iterations needed to obtain convergence increases dramatically when the dimension is huge since each iteration of the algorithm initially requires accessing all the data. For this reason, researchers use one of the following approximation methods.

B. Variational inference (VI)

An approximation technique known as "variational inference" builds the posterior distribution over BNN weights. The Bayesian inference problem is viewed by VI-based techniques as an optimization problem that is utilized to train deep neural networks by the stochastic gradient descant (SGD) Figure 6 summarizes different VI techniques for BNNs [50].

For Bayesian neural networks, the aim of VI-based techniques is to approximation posterior distributions on the weights of the neural networks [45]. To attain this can be described the loss function as follows:

$$\mathcal{L}(\Phi) \approx \frac{1}{2|D|} \sum_{i=1}^{|D|} \mathcal{L}_R\big(y^{(i)}, \mathbf{x}^{(i)}\big) + \frac{1}{|D|} KL\big(q_\phi(w) \parallel p(w)\big), \tag{7}$$

where |D| denotes the number of samples, and *KL* is a Kullback Leibler divergence, which may be used to measure the difference between two probability distributions.

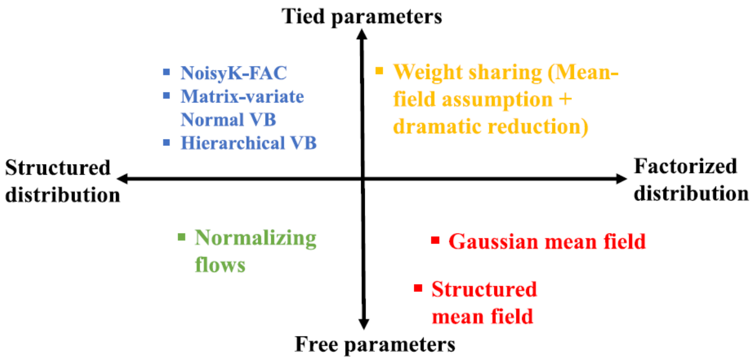


Figure 6. Summarize different variational inference approaches for BDL. Remark that weight sharing (mean-field assumption+dramatic reduction) was added based on the approach proposed in [50]

C. MC-dropout

MC dropout is a prevalent "free lunch" technique in medical imaging for approximate Bayesian computations (ABC). Its appeal is to uncertainty quantification in NNs and solves the challenging problems of ABC, falls within the VI framework; and suggests a multimodal accurate predictive posterior. Estimation of the posterior distribution is the major drawback of the Bayesian networks. It is intractable, in most scenarios. However, Gal and Ghahramani [51] displayed that numerous forward passes can be made while dropout is setting in order to estimate the posterior distribution for a single input (it is noted that, in a frequentist setting, dropout is only applied in training time to beat over-fitting). Dropout-induced randomness can assist us in approximating the posterior distribution with the least amount of statistical difficulty.

3. LITERATURE REVIEW

In recent, several studies on BDL and uncertainty awareness have been performed in models that are applied to the different public datasets for diabetic retinopathy. For example, Garifullin *et al.* [52] proposed a Bayesian baseline for the DR lesion segmentation that allows the model calibration, analysis of segmentation distributions and prediction uncertainty. Furthermore, the challenges regarding the uncertainty quantification and deep probabilistic model are presented. The weak point and problem of this work is the low sensitivity of the model segmentation, they experimented with many loss functions for this problem. However, they did not achieve the required results for the proposed baseline method, which achieved the PR-AUC of 0.483 for microaneurysms, 0.84 for hard exudates, 0.593 for hemorrhages and 0.641 for soft exudates on a dataset of IDRiD. In another study by Jaskari *et al.* [53], they explored the benefits of estimating uncertainty for a clinical dataset of Finland hospital with other publicly available datasets. They applied several different approximations of Bayesian methods that widely used recently. In addition, they proposed a novel uncertainty measure risk-based classier. This improves distribution uncertainty performance on the EyePACS dataset and the Finland Hospital dataset. The strength of this study comes from applying Bayesian approximate methods on different datasets and comparing their performance.

Krishnan *et al.* [54] proposed the MOPED method which is (model priors with empirical Bayes using DNN) for selecting informed weight prior comprising two stages. Firstly, find the maximum-likelihood estimates for weights by using DNN and then setting up the weight prior using empirical Bayes technique to conclude the posterior with VI. The strength of this paper emerges when the proposed method was compared with deep model ensembles and another Bayesian method by using multiple different DL architectures and medical and non-medical datasets. The suggested method demonstrates superior improvement in the performance for accuracy of 0.937 and AUC of 0.912. Following the previous study, the same authors Krishnan *et al.* [55] evaluated the recently proposed MOPED method for BDL benchmarking framework. They have benchmarked MOPED with mean-field VI on a real-world DR diagnosis task and compared it with the latest BDL methods. They showed that the MOPED method provides reliable uncertainty estimations while outperforming the latest methods, offering a novel strong baseline for the community of BDL to contrast complex real-world tasks including larger models.

Singh *et al.* [56] conducted an uncertainty analysis of a DL model for the diagnosis of 4 retinal diseases, using images from a publicly available OCTID dataset. At the test time, MC-dropout was employed to build parameters distribution, and the predictions approximated the predictive posterior of a Bayesian model. According to their information, explained this is one of the first studies of its sort and the use of both uncertainty and explanation has effects on accepting DL models. Both artificial intelligence researchers and the clinical community will benefit from this, which is considered the paper's strengths point. In another work, Band *et al.* [57] presented DR detection benchmarking tasks for Bayesian deep learning. They evaluated well-established and state-of-the-art non-Bayesian and Bayesian methods on a group of task-specific performance and reliability metrics. This study's strength comes from using several Bayesian inference methods (MFVI, SFVI, MC-dropout, MAP and deep ensemble) applied to two large types of retinal datasets EyePACS and APTOS. Also provided implementations of all benchmark methods, as well as results calculated over 20 GPU days, 100 TPU days, 400 hyper-parameters configurations and at least an evaluation on 6 random seeds for each.

Ahsan *et al.* [58] proposed a hybrid model for the problem of DR classification that jointly handles uncertainty problem and is also able to learn from unlabeled data. In particular, their proposed framework has two main components: the Bayesian convolutional neural network (BCNN) model having Monte-Carlo dropout, which is used as a feature descriptor, and an active learning (AL) component. BCNN reduced the uncertainty of the prediction, while the AL module enables learning from unlabeled data. They concluded that their proposed model outperformed state-of-the-art performance in terms of different metrics. In another study by Akbar and Midhunchakkaravarthy [59] designed a novel filtered-based segmentation framework and implemented on the large diabetic retinopathy feature space. In this work, they designed and implemented the proposed model in several phase. First, for fundus training to detection the shape and statistical analysis for classify disease severity, a new mathematical filter used, second to find essential features for problem of classification they applied deep CNN on filtered data. In DL phase, to filter these essential features in each training image for problem of classification the C3D pre trained framework used. At last, in classification phase, to train the features of C3D for disease class prediction a nonlinear hybrid Bayesian SVM classifier is used. Experimental results confirmed that the suggested filtered segmentation-based Bayesian DNN has better runtime and accuracy than the conventional models on different DR datasets. Despite these strengths point, details of Bayesian method that used were not given.

Kwon *et al.* [60] proposed a new way of quantification uncertainty in classification using Bayesian neural network models. Presented the proposed method using two medical datasets, ISLES and DRIVE. The proposed method applied variational inference as a Bayesian approximate method for uncertainty quantification. The suggested method has some advantages over the existing method in that it is numerically stable and expresses the inherent variability in terms of the outcome's underlying distribution. In Ayhan *et al.* [61] presented an intuitive framework based on the augmentation of test time data for quantifying the uncertainty of the latest DNN for DR diagnosing applied (TTAUG) to the case of DR detection, a diagnostic task that is well under 80 stood, for which demonstrated high-performance of DNNs. The proposed framework used a non-Bayesian ensemble approach to quantify uncertainty in DNNs. The results obtained showed that the derivative uncertainty scale is okay-calibrated and experienced clinicians also discover cases with uncertain diagnoses problematic to assess. Singh [62] applied a Bayesian method to optimize and implement different deep learning models to complete DL with Bayesian thought. The proposed BDL approach outperformed existing methods of prediction and diagnosis and demonstrated high precision. With over 98% accuracy, one prototype and two data sets were used for predictions and clinical diagnoses for diabetes and cancer. A weakness of this paper is that no details are given about the Bayesian methods used and no details of datasets that were used.

Toledo-Cortés *et al.* [63] introduced a hybrid deep learning-Gaussian process method for binary classification tasks and uncertainty quantification. The proposed method used Radial Basis Function as an approximation for uncertainty quantification. The strength of this study comes from applying the DLGP-DR method in two different datasets Messidor-2 and EyePACS for DR and comparing the performance of the proposed method with other studies. According to the authors, their results outperformed other studies, which demonstrated that quantification of uncertainty in the prediction improves its interpretability as a diagnostic support tool by enhancing its interpretation. In another study by Filos *et al.* [64], a new benchmark for the deep Bayesian models is presented using an application of real-world medical imaging in diagnosing DR. They applied different Bayesian techniques such as MC-dropout and MFVI with two different datasets (APTOS and EyePACS) that commonly used for DR. Comparing the performance of the methods used, it appears that MC-dropout outperformed other methods.

Farquhar *et al.* [65] proposed the Radial Bayesian neural network that identifies a simple approximates posterior distribution in a hyper spherical space. The proposed method avoids a sampling problem in MFVI brought on by multivariate Gaussians' so-called "soap-bubble" disease. The strong point of this study is that the radial Bayesian neural network outperformed by a wide margin MFVI, and even outperform deep ensemble and MC-dropout models. Toledo-Cortés *et al.* [66] presented a Deep Probabilistic Learning Ordinal Regression model for the diagnosis of medical images. An evaluation of the method was conducted on two various medical image analysis tasks: the diagnosis of prostate cancer and the estimation of diabetic retinopathy grade on the fundus image. Used the radial basis function as an approximation for uncertainty quantification. They provided the details of all methods used in the two tasks. The suggested method Improves the performance of diagnosis on both tasks, as well as interpretability of outcomes by quantifying prediction uncertainty compared to regression architectures and conventional deep classification. These considered the strengths point of the paper. Lim *et al.* [67] proposed to use stochastic batch normalization to compute uncertainty estimations of the DL system’s prediction. They evaluated the effectiveness of utilizing such estimations in a real-life application for the checking of DR.

4. DISCUSSION

Table 3 illustrates a summary of the studies reviewed that applied BDL methods to diabetic retinopathy disease. As we noted in Table 3, different Bayesian techniques were used for different purposes with different model structures. For example, in [60] the VI and variational dropout technique were used for the classification of the (DRIVE) retinal image dataset. Their findings showed that the proposed uncertainty quantification method can be extended to the different functional shapes of variational densities. It can be used for general Bayesian learning using MCMC sampling method, rather than VI as an approximation method. Similarly, [52] used variational dropout but for DR image segmentation using Dense-FCN. In addition, [55] and [58] used the MC dropout technique, while in [55] it is used to diagnose retinal diseases, including diabetic retinopathy. Their results showed that the uncertainty threshold was successful in improving the performance of the model. The use of both uncertainty and interpretability has implications in deep learning models acceptance. However, used BDL for binary and multi class classification images based on CNN model [58]. Their model for binary and multi class classification achieved an accuracy of 92%, and they suggested using other methods instead of MC-dropout such as variational inference. Such methods approximate the posterior-distribution by reducing the KL-divergence between the two distributions that can be investigated for DR classification task, especially in multi classification.

Furthermore, used several Bayesian techniques like MC dropout, MFVI, deep ensemble and GVI for classification of DR using the CNN model [53]. Whereas, they used another method which is radial BNN and radial ensemble using VGG-16 architecture for classification [65]. They identified an issue with sampling during MFVI failures, and they showed that their Radial BNN performs better than MFVI and even MC dropout by a significant margin. Furthermore, used several methods such as (MAP, FSVI, MC-dropout, deep ensemble, MFVI and RADIALMFVI) for DR detection benchmarking tasks for Bayesian deep learning using ResNet architecture [57]. Used MOPED MFVI in different datasets including DR dataset for diagnoses of DR using several architectures such as VGG, ResNet, SCNN and LeNet [56]. Their results provided support for the proposed approach, and provided better performance of the model and reliable estimates of uncertainty in real-world tasks with large-scale complex models. Finally, we noted that the use of BDL is i[ncreasing, especially with the medical images because it can obtain results with high accuracy while taking the model uncertainty into consideration. We also noticed that MCMC was not used in the reviewed studies, and the most common methods used were MC-dropout and MFVI. This is because while MCMC is quite effective in Bayesian inference, it is time consuming compared to other approximation methods, particularly for DL models with thousands or even millions of parameters.

Table 3. Summary of some recent work that used BDL for diabetic retinopathy

Reference	Year	Datasets used	Bayesian technique	Purpose	DL Model
Filos *et al.* [64]	2019	EyePACS from Kaggle APTOS	MC-dropout, MFVI	Classification	VGG
Krishnan *et al.* [55]	2019	EyePACS from Kaggle	MOPED-MFVI	Diagnosis of DR	VGG
Band *et al.* [57]	2020	APTOS and EyePACS From Kaggle	MAP, FSVI, MC-dropout, Deep ensemble, MFVI and RADIALMFVI	Classification	ResNet-50
Akbar and Midhunchakkaravarthy [59]	2020	EyePACS from Kaggle	Unknown	Image segmentation	CNN,C3D and BSVM
Kwon *et al.* [60]	2020	DRIVE	VI and variational dropout	Classification	CNN
Farquhar *et al.* [65]	2020	EyePACS from Kaggle	Radial BNN and Radial Ensemble	Classification	VGG-16
Ahsan *et al.* [58]	2020	APTOS	MC dropout	Binary and multi classification of DR	CNN
Ayhan *et al.* [61]	2020	EyePACS from Kaggle and IDRiD	Ensemble	Detection of DR	ResNet
Toledo-Cortés *et al.* [63]	2020	EyePACS from Kaggle and Messidor-2	Radial Basis Function (RBF)	Classification	DL-GP using CNN
Krishnan *et al.* [54]	2020	EyePACS from Kaggle	MOPED_MFVI	Diagnosis of DR	Resnet-20, SCNN LeNeT, VGG for DR
Toledo-Cortés *et al.* [66]	2022	EyePACS from Kaggle	RBF	Classification	DQOR using CNN
Garifullin *et al.* [52]	2021	IDRiD	Variational dropout	Image Segmentation	Dense-FCN
Singh *et al.* [56]	2021	OCTID	MC dropout	Diagnosis DR	CNN
Jaskari *et al.* [53]	2022	EyePACS, KSSHP, Messidor-2 and APTOS	MC dropout, MFVI, Deep ensemble, GVI and radial BNN	Classification	CNN

5. CONCLUSION

Globally, DR is one of the major causes of blindness, and early detection of DR and timely treatment is critical to reduce the incidence of loss vision. DL methods provides better performance in early detection of DR, and they have been used widely in DR detection and classification. However, despite their widespread use, classical DL models are incapable to quantify the uncertainty in the models, thus they are prone to make wrong decisions in complex cases. Therefore, in recent years, Bayesian DL were developed to provides an accurate framework to identify all uncertainties in the model. This paper reviewed the most published work that applied BDL methods in DR disease. In addition, the most popular Bayesian inference methods in DL were presented, such us MC-dropout, MCMC and variational Inference. We conclude that the use of BDL is increasing, especially with medical images. This is because BDL can obtain results with high accuracy, while quantifying the model uncertainty. Thus, BDL are very useful due to its ability to deal with the two types of uncertainty: (Aleatoric and Epistemic). In the future, we will apply BDL to other medical problems using different prior distributions.

REFERENCE

[1] M. Tan and Q. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, May 2019, vol. 97, pp. 6105–6114.

[2] T. Brown *et al.*, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[4] 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*, Jul. 2017, pp. 1321–1330.

[5] H. D. Kabir, A. Khosravi, M. A. Hosen, and S. Nahavandi, "Neural network-based uncertainty quantification: a survey of methodologies and applications," *IEEE Access*, vol. 6, pp. 36218–36234, 2018, doi: 10.1109/ACCESS.2018.2836917.

[6] D. E. Worrall, C. M. Wilson, and G. J. Brostow, "Automated retinopathy of prematurity case detection with convolutional neural networks," in *Deep learning and data labeling for medical applications*, Springer, 2016, pp. 68–76, doi: 10.1007/978-3-319-46976-8_8.

[7] L. P. Levasseur, Y. D. Hezaveh, and R. H. Wechsler, "Uncertainties in parameters estimated with neural networks: application to strong gravitational lensing," *Astrophys. J.*, vol. 850, no. 1, p. L7, Nov. 2017, doi: 10.3847/2041-8213/aa9704.

[8] A. Kendall, V. Badrinarayanan, and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," 2015, *arXiv: 151102680*.

[9] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Advances in Neural Information Processing Systems*, 2017, vol. 30.

[10] D. S. W. Ting *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, Dec. 2017, doi: 10.1001/jama.2017.18152.

[11] J. F. Arévalo, A. F. Lasave, D. G. Zeballos, and S. Bonafonte-Royo, "Diabetic retinopathy," in *Retinal and Choroidal Manifestations of Selected Systemic Diseases*, J. F. Arévalo, Ed. New York, NY: Springer, 2013, pp. 387–416, doi: 10.1007/978-1-4614-3646-1_21.

[12] C. P. Wilkinson *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, Sep. 2003, doi: 10.1016/S0161-6420(03)00475-5.

[13] C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl, "Leveraging uncertainty information from deep neural networks for disease detection," *Sci. Rep.*, vol. 7, no. 1, p. 17816, Dec. 2017, doi: 10.1038/s41598-017-17876-z.

[14] S. Farquhar, M. A. Osborne, and Y. Gal, "Radial Bayesian neural networks: beyond discrete support in large-scale Bayesian deep learning," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Jun. 2020, pp. 1352–1362.

[15] P. Ruamviboonsuk *et al.*, "Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program," *Npj Digit. Med.*, vol. 2, no. 1, Art. no. 1, Apr. 2019, doi: 10.1038/s41746-019-0099-8.

[16] J. Mukhoti and Y. Gal, "Evaluating Bayesian deep learning methods for semantic segmentation," 2019, *arXiv: 1811.12709*.

[17] L. V. Jospin, W. Buntine, F. Boussaid, H. Laga, and M. Bennamoun, "Hands-on Bayesian neural networks -- a tutorial for deep learning users," 2022, *arXiv: 2007.06823*.

[18] "Diabetic Retinopathy Detection," https://kaggle.com/competitions/diabetic-retinopathy-detection (accessed Jun. 16, 2022).

[19] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Inf. Sci.*, vol. 501, pp. 511–522, Oct. 2019, doi: 10.1016/j.ins.2019.06.011.

[20] T. Kauppi *et al.*, "DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms," *Mach. Vis. Pattern Recognit. Res. Group Lappeenranta Univ. Technol. Finl.*, vol. 73, pp. 1–17, 2006.

[21] "Shiley Eye Institute | UC San Diego." https://shileyeye.ucsd.edu/ (accessed Jun. 16, 2022).

[22] E. Decencière *et al.*, "Feedback on a publicly distributed image database: the Messidor database," *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, 2014.

[23] S. Pachade *et al.*, "Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research," *Data*, vol. 6, no. 2, Art. no. 2, Feb. 2021, doi: 10.3390/data6020014.

[24] "APTOS 2019 Blindness Detection." https://kaggle.com/competitions/aptos2019-blindness-detection (accessed Jun. 16, 2022).

[25] L. Giancardo *et al.*, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Medical Image Analysis*, vol. 16, no. 1, pp. 216-226, Jan. 2012, doi: 10.1016/j.media.2011.07.004.

[26] E. Decencière *et al.*, "TeleOphta: machine learning and image processing methods for teleophthalmology," *IRBM*, vol. 34, no. 2, pp. 196–203, Apr. 2013, doi: 10.1016/j.irbm.2013.01.010.

[27] M. Niemeijer *et al.*, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010, doi: 10.1109/TMI.2009.2033909.

[28] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. V. Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004, doi: 10.1109/TMI.2004.825627.
[29] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep learning techniques for diabetic retinopathy classification: a survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.
[30] S. Shekar, N. Satpute, and A. Gupta, "Review on diabetic retinopathy with deep learning methods," *J. Med. Imaging*, vol. 8, p. 32, 2021.
[31] "Ocular disease recognition," https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k (accessed Jun. 16, 2022).
[32] P. Porwal *et al.*, "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research," *Data*, vol. 3, no. 3, p. 25, 2018.
[33] T. Kauppi *et al.*, "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *BMVC*, 2007, doi: 10.5244/C.21.15.
[34] B. Tymchenko, P. Marchenko, and D. Spodarets, "Deep learning approach to diabetic retinopathy detection," 2020, *arXiv: 2003.02261.*
[35] A. R. Mary, and M. F. Begam, "A comprehensive survey on deep learning based automated diabetic retinopathy detection," *J. Xidian Univ.*, vol. 14, no. 7, Aug. 2020, doi: 10.37896/jxu14.7/231.
[36] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Comput. Sci.*, vol. 90, pp. 200–205, 2016.
[37] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media, Inc., 2019.
[38] F. Chollet, *Deep learning with Python*. Simon and Schuster, 2021.
[39] G. Zhao, F. Liu, J. A. Oler, M. E. Meyerand, N. H. Kalin, and R. M. Birn, "Bayesian convolutional neural network based MRI brain extraction on nonhuman primates," *NeuroImage*, vol. 175, pp. 32–44, Jul. 2018, doi: 10.1016/j.neuroimage.2018.03.065.
[40] Y. Tang and X. Wu, "Text-independent writer identification via CNN features and joint Bayesian," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Oct. 2016, pp. 566–571, doi: 10.1109/ICFHR.2016.0109.
[41] P. Waldmann, "Approximate Bayesian neural networks in genomic prediction," *Genet. Sel. Evol.*, vol. 50, no. 1, p. 70, Dec. 2018, doi: 10.1186/s12711-018-0439-1.
[42] A. A. Abdullah, M. M. Hassan, and Y. T. Mustafa, "A review on bayesian deep learning in healthcare: applications and challenges," *IEEE Access*, vol. 10, pp. 36538–36562, 2022, doi: 10.1109/ACCESS.2022.3163384.
[43] J. Xuan, J. Lu, and G. Zhang, "A survey on bayesian nonparametric learning," *ACM Comput. Surv.*, vol. 52, no. 1, p. 13:1-13:36, Jan. 2019, doi: 10.1145/3291044.
[44] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, Mar. 2021, doi: 10.1007/s10994-021-05946-3.
[45] M. Abdar *et al.,* "A review of uncertainty quantification in deep learning: techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243-297, 2021.
[46] H. Wang and D.-Y. Yeung, "A survey on bayesian deep learning," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–37, Oct. 2020, doi: 10.1145/3409383.
[47] T. Holovko, "Quantifying uncertainty in deep learning," *PhD diss.*, 2020.
[48] M. A. Kupinski, J. W. Hoppin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," *JOSA A*, vol. 20, no. 3, pp. 430–438, 2003.
[49] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning*, 2016, pp. 1050-1059.
[50] J. Swiatkowski *et al.*, "The k-tied normal distribution: a compact parameterization of gaussian mean field posteriors in Bayesian neural networks," 2020, *arXiv: 2002.02655*.
[51] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: representing model uncertainty in deep learning," in *Proceedings of The 33rd International Conference on Machine Learning*, Jun. 2016, pp. 1050–1059.
[52] A. Garifullin, L. Lensu, and H. Uusitalo, "Deep Bayesian baseline for segmenting diabetic retinopathy lesions: Advances and challenges," *Comput. Biol. Med.*, vol. 136, p. 104725, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104725.
[53] J. Jaskari *et al.*, "Uncertainty-aware deep learning methods for robust diabetic retinopathy classification," 2022, *arXiv:* 2201.09042.
[54] R. Krishnan, M. Subedar, and O. Tickoo, "Specifying Weight priors in bayesian deep neural networks with empirical Bayes," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, Apr. 2020, pp. 4477–4484, doi: 10.1609/aaai.v34i04.5875.
[55] R. Krishnan, M. Subedar, O. Tickoo, A. Filos, and Y. Gal, "Improving MFVI in Bayesian neural networks with empirical Bayes: a study with diabetic retinopathy diagnosis," *Fourth workshop on Bayesian Deep Learning (NeurIPS 2019)*, pp. 1-7, 2019.
[56] A. Singh, S. Sengupta, A. Rasheed, V. Jayakumar, and V. Lakshminarayanan, "Uncertainty aware and explainable diagnosis of retinal disease," *In Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11601, pp. 116-125, 2021.
[57] N. Band *et al.*, "Benchmarking Bayesian deep learning on diabetic retinopathy detection tasks," 2022, *arXiv: 2211.12717* 2022.
[58] M. A. Ahsan, A. Qayyum, J. Qadir, and A. Razi, "An active learning method for diabetic retinopathy classification with uncertainty quantification," 2020, *arXiv:* 2012.13325.
[59] S. Akbar and D. Midhunchakkaravarthy, "A novel filtered segmentation-based Bayesian deep neural network framework on large diabetic retinopathy databases," *Rev. Intell. Artif.*, vol. 34, pp. 683–692, Dec. 2020, doi: 10.18280/ria.340602.
[60] Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, "Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation," *Comput. Stat. Data Anal.*, vol. 142, p. 106816, Feb. 2020, doi: 10.1016/j.csda.2019.106816.
[61] M. S. Ayhan, L. Kühlewein, G. Aliyeva, W. Inhoffen, F. Ziemssen, and P. Berens, "Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection," *Med. Image Anal.*, vol. 64, p. 101724, Aug. 2020, doi: 10.1016/j.media.2020.101724.
[62] P. N. Singh, "Better application of bayesian deep learning to diagnose disease," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, Apr. 2021, pp. 928–934, doi: 10.1109/ICCMC51019.2021.9418301.
[63] S. Toledo-Cortés, M. d. l. Pava, O. Perdomo, and F. A. González, "Hybrid deep learning gaussian process for diabetic retinopathy diagnosis and uncertainty quantification," in *Ophthalmic Medical Image Analysis*, 2020, pp. 206–215. doi: 10.1007/978-3-030-63419-3_21.
[64] A. Filos *et al.*, "A systematic comparison of bayesian deep learning robustness in diabetic retinopathy tasks," 2019, *arXiv:* 1912.10481.
[65] S. Farquhar, M. A. Osborne, and Y. Gal, "Radial Bayesian neural networks: beyond discrete support in large-scale bayesian deep learning," *In International Conference on Artificial Intelligence and Statistics,* pp. 1352-1362, 2020.

[66] S. Toledo-Cortés, D. H. Useche, H. Müller, and F. A. González, "Grading diabetic retinopathy and prostate cancer diagnostic images with deep quantum ordinal regression," *Comput. Biol. Med.*, vol. 145, p. 105472, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105472.
[67] Z. Lim, M. Lee, W. Hsu, and T. Y. Wong, "Building trust in deep learning system towards automated disease detection," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 9516–9521, doi: 10.1609/aaai.v33i01.33019516.

BIOGRAPHIES OF AUTHORS

Halbast Rashid Ismael is working as a programmer at the Technical College of Informatics-Akre, Duhok Polytechnic University, Kurdistan Region, Iraq. She was born in Akre, Kurdistan Region, Iraq, in 1987. She got her diploma degree in Computer Systems from Akre Technical Institute in 2008. And bachelor's degree in Information Technology, from the Technical College of Informatics-Akre, Duhok Polytechnic University, Kurdistan Region, Iraq, in 2017. Now she is completing her master's degree (MSc) at Duhok Polytechnic University, Department of Information Technology, Duhok, Iraq. Her research interests include machine learning, artificial intelligence, and the internet of things (IoT). She can be contacted at email: halbast.rashid@dpu.edu.krd.

Masoud Muhammed Hassan is Assistant Professor at Faculty of Science, Computer Science Department, University of Zakho, Kurdistan Region, Iraq. He was born in Zakho, Kurdistan Region, Iraq, in 1982. He received the B.Sc. degree in statistics and informatics from Salahaddin University, in 2005, the M.Sc. degree from the School of Mathematics and Computer Science, University of Mosul, Iraq, in 2008, and the Ph.D. degree in mathematics and statistics from The University of Sheffield, Sheffield, U.K., in 2016. He is currently working as a Resesarcr Lecturer at the Department of Computer Science, University of Zakho. His research interests include statistical modeling, machine learning, data mining, and artificial intelligence. He can be contacted at email: masoud.hassan@uoz.edu.krd.