

Received March 2, 2022, accepted March 25, 2022, date of publication March 30, 2022, date of current version April 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3163384

A Review on Bayesian Deep Learning in Healthcare: Applications and Challenges

ABDULLAH A. ABDULLAH¹, MASOUD M. HASSAN^{1,2},
AND YASEEN T. MUSTAFA^{3,4}, (Member, IEEE)

¹Computer Science Department, Faculty of Science, University of Zakho, Duhok 42002, Iraq

²College of Arts and Sciences, The American University of Kurdistan, Duhok 44001, Iraq

³Computer Science Department, College of Science, Nawroz University, Duhok 44001, Iraq

⁴Environmental Science Department, Faculty of Science, University of Zakho, Duhok 42002, Iraq

Corresponding author: Abdullah A. Abdullah (abdullah.abdullah@uoz.edu.krd)

ABSTRACT In the last decade, Deep Learning (DL) has revolutionized the use of artificial intelligence, and it has been deployed in different fields of healthcare applications such as image processing, natural language processing, and signal processing. DL models have also been intensely used in different tasks of healthcare such as disease diagnostics and treatments. Deep learning techniques have surpassed other machine learning algorithms and proved to be the ultimate tools for many state-of-the-art applications. Despite all that success, classical deep learning has limitations and their models tend to be very confident about their predicted decisions because it does not know when it makes mistake. For the healthcare field, this limitation can have a negative impact on models predictions since almost all decisions regarding patients and diseases are sensitive. Therefore, Bayesian deep learning (BDL) has been developed to overcome these limitations. Unlike classical DL, BDL uses probability distributions for the model parameters, which makes it possible to estimate the whole uncertainties associated with the predicted outputs. In this regard, BDL offers a rigorous framework to quantify all sources of uncertainties in the model. This study reviews popular techniques of using Bayesian deep learning with their benefits and limitations. It also reviewed recent deep learning architecture such as Convolutional Neural Networks and Recurrent Neural Networks. In particular, the applications of Bayesian deep learning in healthcare have been discussed such as its use in medical imaging tasks, clinical signal processing, medical natural language processing, and electronic health records. Furthermore, this paper has covered the deployment of Bayesian deep learning for some of the widespread diseases. This paper has also discussed the fundamental research challenges and highlighted some research gaps in both the Bayesian deep learning and healthcare perspective.

INDEX TERMS Bayesian deep learning, Bayesian neural networks, deep learning, healthcare, MCMC, MC-dropout, variational inference.

I. INTRODUCTION

Machine learning has grown in popularity in the last decade, expanding from applications of models from a few available datasets to a wide range of scientific and technological fields. Deep learning (DL) is a subfield of machine learning which employs neural network structures that consist of an input layer, an output layer, and multiple hidden layers that can be range from two to tens or hundreds of layers with millions or even billions number of parameters, such as ResNetv2 and GPT-3 [1]. The ability to extract hidden useful knowledge

in data as well as the availability of a large amount of data made DL models more successful as they are faster at processing a big amount of data, and this made DL make a historical shift from classical machine learning techniques. This success has given DL the opportunity to be applied in a variety of scientific domains, among which is healthcare. DL in healthcare has advanced to the point that it can now be applied in several subfields that have proven to outperform human skills, such as medical imaging [2]. These models can take different types of data as input, such as images, text, and health records, and produce various types of output, such as generating images, predicting classes, and analyzing text, with the ability to handle multiple different types of inputs

The associate editor coordinating the review of this manuscript and approving it for publication was Felix Albu¹.

at once and generate multiple types of output in complex models. In addition to medical imaging, deep learning methods have been extensively investigated in disease diagnostics, pharmaceuticals, and drug discovery [3].

Despite its promise and advancement, which may surpass medical practitioners in many circumstances, deep learning has only been used sparingly in healthcare disciplines such as medical imaging readings. However, as compared to a decade ago, the number of DL applications in use has increased dramatically [4], [5]. Moreover, emerging technologies such as cloud computing, big data processing have contributed to advancing deep learning techniques in healthcare [6], [7].

Despite the aforementioned potential and additional benefits of deep learning, classical deep learning methods suffer from overfitting and their models are data-hungry. Many strategies have been created to tackle the overfitting problem, the bulk of which were developed in the recent decade, such as dropout, Lasso Regression (L1)/ Ridge Regression (L2) regularizations, Drop-connect, etc [8]. These strategies reinforced deep learning models, allowing them to overcome the overfitting issue and to be less vulnerable to bias and noise associated with data. Yet, classical deep learning models appear to be perfect tools for knowledge extraction from data; but is this the case? To answer this question the output of these models has to be evaluated. Classical deep learning models have promising results compared to other machine learning algorithms. However, such models use the maximum likelihood approach to update model parameters and all of the parameters used in these models are a single-point estimate. Yet, like other machine learning algorithms, classical deep learning models are too confident about their output and decision. This is due to the fact that these models are uncertain about their outputs and the model “doesn’t know when it doesn’t know” [8]. Thus, classical deep learning models are naively confident without taking into consideration the uncertainty associated with data and the model itself. Exploring and quantifying such uncertainties is very crucial for deep learning models, particularly in decision-sensitive contexts such as healthcare applications. To achieve this, Bayesian neural networks emerged and gained popularity in recent years. Bayesian neural networks, also known as Bayesian deep learning (BDL), use probability theory techniques to extract knowledge from provided datasets. This is performed by combining prior information and the likelihood of data to provide posterior distributions, and thus make inferences about the model’s unknown parameters while also dealing with model uncertainty. These two names are being used interchangeably in the research community, for sake of simplicity, Bayesian deep learning (BDL) will be used throughout this work. BDL has better performance, particularly where data are insufficient or scarce, especially when it is hard or expensive to get more data. Despite the fact that BDL outperforms classical deep learning in most cases, it is underused by the research community and has a long way ahead before it reaches its full potential.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, this is the first review effort that addresses the usage of BDL in healthcare to benefit the research community.
- Reviewing the Bayesian methods for deep learning models.
- A comprehensive review of BDL in healthcare applications, such as disease diagnostics/ detection, medical imaging, clinical signal processing, and electronic health records, is provided.
- Main challenges in both BDL and healthcare prospective, are highlighted.
- Some research gaps and open issues in the field that need further investigation are discussed.

The rest of this work is organized as follows. Section 2 discusses the main research gaps. Section 3 introduces Bayesian deep learning. Section 4 briefly overviews the deep learning models and architectures. A review of the most popular Bayesian techniques for deep learning models is provided in section 5. Section 6 covers a review of BDL applications in healthcare, such as medical imaging tasks, signal processing, natural language processing, electronic health record, and audio processing. Discussion of BDL application on diseases diagnostics covered in section 7. Challenges facing the implementation of BDL in healthcare from the BDL and healthcare perspective are highlighted in section 8. In addition, some open issues and research gaps are discussed in the same section. Finally, section 9 concludes the remarks of this work.

II. RESEARCH GAP AND MOTIVATION

The existence of, object detection, and so on. The existence of research covering different aspects of BDL in healthcare is essential for assisting newcomers in the field in identifying knowledge gaps. In classical deep learning, Shamsirband *et al.* [9] have published a review on classical deep learning in healthcare in 2020. They discussed deep learning architectures that are commonly used in healthcare, DL models for specific disorders, and a comparison of DL approaches in healthcare. Nisar *et al.* in [10] published a review paper in 2021, which covers different aspects of classical DL in healthcare. Flowing their introduction, the paper covered the following topics: (1) some deep learning methods and their use in healthcare, (2) DL use in the nervous system, cardiovascular system, and respiratory system, (3) methods, datasets, and applications of DL in healthcare, (4) Some research opportunities and challenges of DL in healthcare. Qayyum *et al.* [11] presented a study in 2020 in which they reviewed classical DL approaches in healthcare with relation to security and privacy considerations, as well as challenges in this regard. In a different section of the paper, the authors presented possible solutions to the concerns discussed in previous sections, as well as a list of research gaps in the field. Pandey *et al.* [12] presented a review article in 2019 that covered the following topics: deep learning

methods and techniques in healthcare, some experimental analysis of DL methods, challenges of DL in healthcare, and some applications of deep learning in the field of healthcare. Kim *et al.* [13] released a paper in 2019 that focused on medical imaging. Many areas of classical DL for images in the realm of healthcare were discussed in the study, including classification, segmentation registration, object detection, and so on.

Few reviews on BDL have been published in recent years. In 2017, Polson *et al.* [14] have published a review paper on BDL. Their review covered the following topics: probabilistic deep learning, how to find a good predictor for Bayes, algorithms for model learning, an example of an application, and potential future research directions in the discussion part. Wang *et al.* [15] have published a review on BDL in 2016 and the same researchers extended the same review and published another paper [16] in 2020. In later work, these topics are covered: DL architectures, probabilistic graphical models, BDL in detail, as well as some models and applications of BDL. In 2019, Xuan *et al.* [17] published a review in this manner and covered some definitions of the field, stochastic process and its manipulation, posterior inference, application for machine learning tasks, and some real-world applications. In 2020, Charnock *et al.* [18] published a paper on BDL that addressed various types of uncertainty, Bayesian neural networks with applicable methods, practical implementations of BDL in two approaches (numerically using MCMC, and approximation methods). In addition to reviews on BDL, few researchers have worked on Uncertainty Quantification (UQ) as BDL has widely contributed to the topic. In 2021, Abdar *et al.* [19] have published an extensive review on UQ in DL. In regards to BDL, the study explored Bayesian techniques for UQ, various Bayesian and other DL applications for machine learning tasks, as well as some research gaps and future directions. Moreover, Alizadehsani *et al.* [20] presented a detailed study on uncertainty handling in medical data, which covered work done in the previous 30 years. The paper also looked at some previous work on Bayesian inference for UQ.

A wide overview that covers most of the topics linked to BDL in healthcare would save time and effort that may be spent reading much-related work to discover the field's research area. Despite the existence of various review publications in the field of classical DL and BDL in healthcare, such as those listed above, there is no review of publicly available work particularly on BDL in healthcare to the best of our knowledge. To overcome the shortage of reviews in BDL in healthcare, this work aims to provide a comprehensive review that covers different aspects of employing BDL in healthcare. This work intends to serve as a starting point for researchers interested in the BDL in healthcare, and materials contained in it are targeted at newcomers to the field as well as BDL researchers seeking research gaps in the healthcare field. This paper covers all the following: Bayesian inference and approximation methods, Bayesian deep learning in healthcare applications, disease diagnostics, and some

challenges and/or opportunities for future research in the field.

III. BAYESIAN DEEP LEARNING

BDL refers to probabilistic deep learning that is based on the Bayes theorem. Bayesian methods combine prior "expert" information and the likelihood of data to produce posterior distributions, which can represent different uncertainties in the model [8]. It is worth mentioning that there is another term used in some of the published work, which is the Bayesian treatment for the neural network "sometimes referred to as the Bayesian neural network." Bayesian treatment for neural networks is a more general topic of using the Bayesian method in neural network models where Bayesian methods were used for tuning weights, using Bayesian optimization inference, using Bayesian activation functions. On the other hand, BDL, "which can also be called a Bayesian neural network," is concerned with estimating the posterior distribution of data, especially to deal with different kinds of uncertainties. BDL has several advantages compared to classical deep learning [17]. First, in addition to dealing with overfitting, particularly when the data is insufficient to feed the model, BDL is used to represent and quantify uncertainties of DL models based on the probabilistic foundations of Bayesian statistics. Second, it makes advantages of using experts' prior knowledge, which refers to beliefs about data and its distribution before seeing any data. Third, BDL uses probability distributions for model parameters "weights and biases", which implicitly interpret model output without the need of doing multiple tests and cross-validation of datasets. The prior and likelihood distributions can take any shape of probabilistic distributions such as Normal, Gamma, Beta, and Cauchy. Figure 1 shows the difference between BDL and classical DL.

BDL can deal with two types of uncertainty in general: Aleatoric and Epistemic uncertainty. Aleatoric uncertainty refers to the uncertainty associated with data, which might be due to the nature of the data, noise, or anomalies that exist among data. Whereas, epistemic uncertainty refers to the uncertainty associated with the model structure and model parameters [21], [22]. This is due to the fact that the aleatoric uncertainty is derived from data, and the model developer or researcher has no control over data, as it cannot be reduced [8]. On the other hand, epistemic uncertainty can be controlled, and the Bayesian method is an efficient way to deal with this type of uncertainty while avoiding overfitting [23]. In contrast to other methods, when the same input is fed to the model, BDL produces different outputs, which is caused by sampling methods and probability distributions used.

The main idea behind the BDL is to define the posterior of data for the neural networks model to quantify the model's uncertainty. To calculate posterior, the prior distribution $P(\omega)$ is chosen for the model parameters, which is primarily based on previous experience, and most of the time the Gaussian/Normal distribution is usually used as the

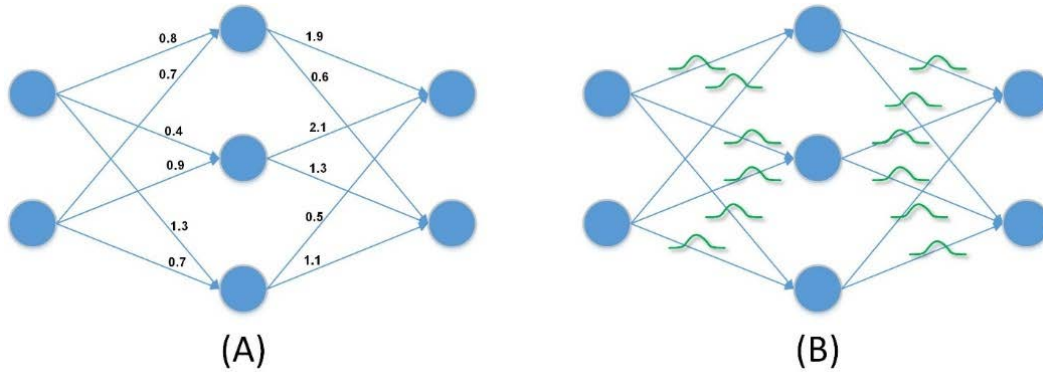


FIGURE 1. Basic Neural Network. (A) a simple classical NN, (B) a simple Bayesian NN.

default prior. The likelihood $P(D|\omega)$ is calculated using given data before calculating the posterior (conditional probability). The posterior is always proportional to the prior times' likelihood. The Bayes theorem is defined in (1).

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)} \quad (1)$$

In BDL, the probability of parameters given data $P(\omega|D)$ is used, and the Bayesian equation can be defined as in (2).

$$P(\omega|D) = \frac{P(\omega) \cdot P(D|\omega)}{P(D)} \quad (2)$$

The term marginal likelihood (evidence) which is $P(D)$ is used to normalize the posterior.

To generate the posterior $P(\omega|D)$ distribution, the product of likelihood and prior are computed for all parameters, which involves finding integrals over all parameters, which is referred to as $P(D)$ or evidence "marginal likelihood". The integral may be calculated analytically if the form of the posterior distribution is known, especially when conjugate probability distributions are used for prior and likelihood. However, in most cases "where distributions are not conjugate", the posterior distribution is intractable, which implies it cannot be solved analytically. The posterior is calculated in this manner using approximation methods, often known as sampling. With all of its variances, Markov Chain Monte Carlo (MCMC) is the most frequently used sampling method for Bayesian inference. Despite the fact that the MCMC provides promising results when compared to other methods, it has not gained popularity in deep learning models. This is due to computation issues, as this method requires thousands, if not millions, of samples to calculate posterior. To overcome the MCMC computation issues, some approximation methods are gaining popularity in the deep learning community, such as Variational Inference (VI) and Monte Carlo Dropout (MC-Dropout).

To find the predictive distribution $P(y^*|x^*, D)$, weights ω of networks are considered as variables. y^* are predicted outcomes, x^* are observations, and D is training data, while $P(\omega|D)$ is the true posterior distribution. The approximate

posterior $q(\omega)$ is used instead of the true posterior in order to find the predictive distribution, as shown in (3).

$$P(y^*|x^*, D) = \int P(y^*|x^*, \omega) P(\omega|D) d\omega \approx \int P(y^*|x^*, \omega) q(\omega) d\omega \quad (3)$$

IV. DEEP LEARNING ALGORITHMS AND ARCHITECTURES

In its basic form, deep learning refers to neural networks with more than one hidden layer between input and output. Current state-of-the-art deep learning models have hundreds of these hidden layers such as in the GPT-3 model [24]. Although there are many distinct layers in deep learning models, the main architectures used in most deep learning models are Fully Connected Neural Networks (FCNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). Deep learning models can have thousands or millions of parameters to be trained, and complex models can have billions of parameters. Model weights (parameters) can be tuned using optimization algorithms such as Gradient Decent (GD), Stochastic Gradient Decent (SGD), Adam, and others, where each parameter in classical deep learning models has a single value. However, for BDL, it is represented by a distribution that takes the number of parameters associated with undertaken distribution. The normal distribution, for example, has mean (μ) and standard deviation (σ) as parameters, for that when upgrading weights of a classical deep learning model to BDL, the number of parameters almost doubled for the same model structure [8].

A. FULLY CONNECTED NEURAL NETWORKS (FCNNs)

The FCNNs are made up of only neurons and layers of models, where all input from previous layers is linked to every neuron in the next layer. The most obvious purpose of neurons is to think about it as linear regression for each neuron, which consists of input data x , weights ω , and bias ω_0 to produce an output of neuron y . The FCNNs are mathematically defined

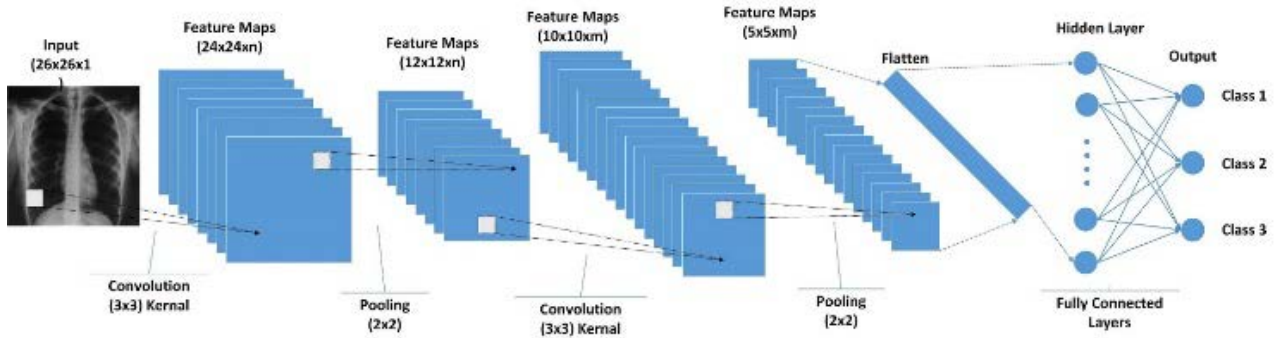


FIGURE 2. A network with CNNs and FC layers.

as seen in (4).

$$y_i = \omega_0 + \sum_{i=1}^n \omega_i x_i \quad (4)$$

The neuron’s value must pass through an activation function, and if the value following the activation function is zero, the neuron will be discarded for that instance [21]. Otherwise, the value will be passed to the next layer in the network. The activation function’s output value varies based on the type of activation function used. The sigmoid activation function, for example, will have an s-shape curve with the output being between 0 and 1. Another example is Rectified Linear Unit (ReLU) activation function. It has a linear shape for values bigger than 0 and a 0 for anything smaller or equal to 0. The mathematical expression of this is shown in (5) and (6).

$$Sigmoid: f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

$$ReLU: f(x) = \max(0, x) \quad (6)$$

Almost every DL models have at least one fully connected instance, even if it is just at the output layer. To update the weights and outputs to probability distributions, almost all early BDL attempts used FCNNs [25], [26]. Figure 3 describes the structure of the FCNNs model.

B. CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNN is the most popular and widely used deep learning architecture for image data. Layers that use convolution methods on some input to produce certain outputs are referred to as CNN. Filters are used as a sliding window in the convolution process to go through all regions of input to produce a feature map. Pooling layers with convolution layers are used in the downsampling process to minimize the amount of produced features and hence the computation. The input can be a tensor of one, two, or three dimensions, and the output takes a similar shape in most cases. Over the years, many CNNs architectures are developed and some of the popular architectures are LaNet, ResNet, VGGNET, EfficientNet, and others [27]. In addition to being the most popular method for image processing, CNNs are also used in other machine learning tasks, such as video processing, natural language

processing NLP, and time series prediction [28]. CNN has also gained popularity in BDL in the last few years, and they have been used in different applications including medical imaging [29], text identification [30], genome studies [31], and others. Figure 2 illustrates the main architecture of the CNNs model.

C. RECURRENT NEURAL NETWORKS (RNNs)

RNNs are one of the most commonly used architecture for deep sequential learning, which is used for sequence or streaming data, such as video, audio, and time-series prediction [32]. These networks contain recurrent “cycle” links between the neurons of the same layer. RNNs contain memory cells, which enable the model to remember data from the past, and that is important for forecasting future outcomes. The RNNs need to keep a state of current and previous input to predict the output of the network. In other words, at least theoretically, the state of all previous inputs is required to compute the result. When the sequence gets bigger, the initial weight is no longer accessible, the vanishing gradient descent problem occurs, and the performance of the RNNs is reduced. Thus, different architectures have been developed to overcome this issue; one of those is Long Short-Term Memory (LSTM) model. The LSTM is recommended for managing each memory cell for both state and output values throughout the learning process. The cell state “C” is the key component of the LSTM model. LSTM does not change cell state or remove its value; it can only regulate it through different gates of LSTM. The forget gate (f_t) decides what information does no longer needed in the cell state based on input (x_t) and previously hidden layer (h_{t-1}) using the sigmoid (σ) function, which works as shown in (7).

$$f_t = \sigma(\omega_f [h_{t-1}, x_t] + b_f) \quad (7)$$

The next gate (layer) in the LSTM decides which information shall be passed or retained in the cell state. It consists of two steps. First, the input gate which takes x_t and h_{t-1} using σ function to calculate the result. Second, the candidate gate \mathbb{C} which takes x_t and h_{t-1} using \tanh function to calculate

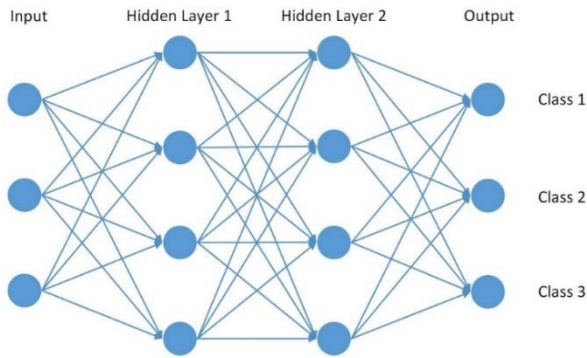


FIGURE 3. A network of fully connected layers.

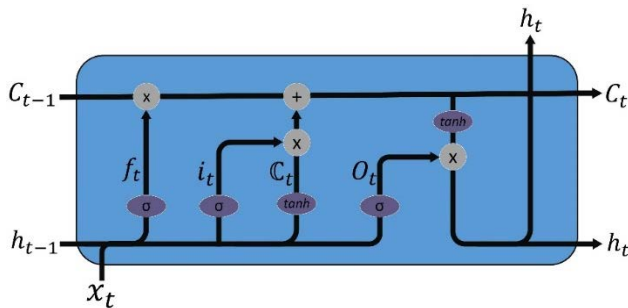


FIGURE 4. LSTM architecture, replica of [114].

results as illustrated in (8) and (9).

$$i_t = \sigma(\omega_i [h_{t-1}, x_t] + b_i) \quad (8)$$

$$C_t = \tanh(\omega_c [h_{t-1}, x_t] + b_c) \quad (9)$$

The current Cell state value is calculated using previous values as indicated in (10).

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (10)$$

Lastly, the output of the LSTM (O_t) is calculated. O_t can be calculated using x_t , and h_{t-1} based on sigmoid function, and multiplied by \tanh of C_t as shown in (11) and (12).

$$O_t = \sigma(\omega_o [h_{t-1}, x_t] + b_o) \quad (11)$$

$$h_t = O_t * \tanh(C_t) \quad (12)$$

In comparison to other architectures of neural networks, RNNs models have not gained popularity in BDL applications. This is due to the computational power needed for complex models. Nonetheless, it has been used by some researchers for different BDL applications, such as Natural Language Processing (NLP) [33] and Forecasting [34]. The LSTM model’s main architecture is depicted in Figure 4.

V. BAYESIAN INFERENCE AND APPROXIMATION METHODS

The learning process in BDL is based on predicting the posterior. It is based on sampling from the posterior distribution using the Bayes theorem, which requires prior and likelihood combinations. However, posterior sampling is a difficult

and computationally expensive process, because of the high dimensionality of the sampling space. In addition, both the prior and the likelihood may belong to different families of distribution that cannot be computed analytically, rendering the posterior untraceable. Despite the need for sampling, only a few sampling methods have gained popularity in BDL models, with the others either requiring further investigation or were not very successful for this task. The most popular sampling algorithm is Markov Chain Monte Carlo (MCMC), which refers to the exact sampling method. The Variational Inference (VI) and MC-Dropout are other sampling methods that are widely used for a posterior estimation, which are considered approximation algorithms. The following are the three sampling methods used to sample from the posterior distribution.

A. MARKOV CHAIN MONTE CARLO (MCMC)

The MCMC sampling method is based on the probability of each generated sample taken into consideration previous samples, resulting in a chain of samples mimicking the target distribution. One property of MCMC sampling is the possibility of transitioning from a given state to a different state within the distribution in finite steps without having to stay in loops in a chain for a long period of time [18]. Samples produced at the beginning of the chain are rejected in this method. This rejection is done because of the error associated with it such as bias toward the initial conditions, and this process of rejection is called burned-in [35]. Although the burned-in wastes some computation and time, it is essential to diminish error in order to achieve better performance to reach the equilibrium distribution of samples. Burned-in sample sizes are not fixed and may require different lengths based on the complexity of the distribution and the MCMC method employed, thus they should be carefully chosen. The large size of burned-in samples requires more time to compute enough samples, while the small size of burned-in may have the effect of the initial condition, thus lowering the performance [35]. Another important issue to consider is the sample correlation because each subsequent sample is calculated using the current sample. To address this issue, some methods of MCMC such as Metropolis Hasting employ step size, which refers to discarding a number of samples before considering a new one [36]. New samples are calculated using (13).

$$\begin{aligned} P(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1) \\ = P(X_t = x_t | X_{t-1} = x_{t-1}) \end{aligned} \quad (13)$$

The MCMC offers the best performance in sampling from the posterior distribution for BDL, but it is computationally costly and difficult to implement [21]. The most popular MCMC methods are Gibbs sampling, Metropolis Hasting, and Hamiltonian Monte Carlo. Gibbs sampling requires knowing the shape of the posterior distribution in order to sample from it. The Metropolis-Hasting algorithm has the advantage of not requiring the true probability distribution to be sampled from; instead, the proportionality of the function

used is sufficient for sampling. The Metropolis-Hasting algorithm begins with the random initial value. Then, the new sample is taken from points near the ascending point. In case the new sample has a better likelihood, the sample is considered. Otherwise, the sample is considered based on a pre-determined probability and rejected if the probability is not met [37]. Hamiltonian Monte Carlo made use of Gibbs sampling walking steps and the Metropolis-Hasting algorithm's acceptance mechanism [18]. Hamiltonian Monte Carlo is the most popular MCMC sampling method for BDL [38], [39].

B. VARIATIONAL INFERENCE (VI)

The MCMC sampling approach is computationally costly and difficult to scale up for complicated deep learning models and large datasets. As a result, various methods for overcoming this constraint have been developed, although with some tradeoff between scalability and performance. In this regard, the variational inference VI is the most popular approximation method used for sampling from the posterior distribution. Despite the fact that VI does not have the same exact distribution; yet, it performs well in estimating the target distribution with considerably less sampling, making it suitable for application in large-scale and complex models and datasets. A number of parameters required for VI to be trained in deep learning models are based on the prior and posterior distributions used. The learning process for models tends to tune these parameters to produce a distribution that is as close to the desired distribution as possible. The Kullback–Leibler (KL) divergence measure can be used to describe the similarity or divergence between two distributions, the posterior “ $p(x_i)$ ” and the variational distribution “ $q(x_i)$ ”, as follows:

$$D_{KL}(p \parallel q) = \sum_{i=1}^n p(x_i) \cdot \log \left(\frac{p(x_i)}{q(x_i)} \right) \quad (14)$$

The goal is to have a variational distribution that is very similar to the posterior, which means that the distance between the variational and true posterior distributions should be reduced [40], [41]. Figure 5 depicts the main difference between MCMC and VI methods.

C. MONTE CARLO DROPOUT (MC-DROPOUT)

Dropout techniques are being used in deep learning models for regularization purposes, which prevent models from overfitting during training [42]. The methodology behind the dropout is shutting down or discarding some neurons in a given layer with a certain probability during the training process for each iteration or epoch. The dropout can be activated during the testing and evaluation process with a certain probability of neuron dropping to estimate the probability distribution of the targeted instance [23]. This technique is referred to as Monte Carlo dropout (MC dropout). It is similar to variational inference used to estimate posterior distribution. This method is even faster than variational inference

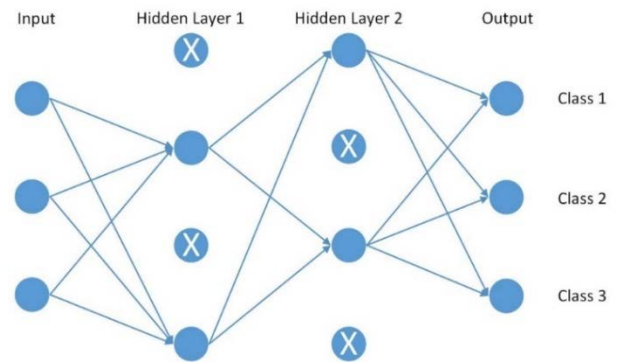


FIGURE 5. MC-Dropout mechanism in DL models.

because it has fewer parameters and requires less time for the model to converge. Figure 6 illustrates the MC-Dropout mechanism.

VI. BDL IN HEALTHCARE SYSTEMS

BDL has been used in healthcare to perform various tasks on various types of data. Despite the fact that it is not widely employed in healthcare, some work has been done covering various aspects of healthcare. In this section, we review the most recent BDL work in the field of healthcare applications.

A. MEDICAL IMAGING

Medical imaging, often known as radiography, is a branch of medicine in which doctors reconstruct images of various body components for diagnostic or therapeutic purposes. BDL has been used in medical imaging to tackle different types of problems associated with it. Compared to other areas, medical imaging gained the most popularity of using BDL. Medical image classification is the most obvious example where DBL is used to classify medical images to improve prediction performance and/or to quantify different types of uncertainties. In addition, there is also work in medical image segmentation, registration, reconstruction, and enhancement.

1) IMAGE CLASSIFICATION

Image classification is the process of arranging images by some given classes to assign labels to each image. This classic image processing task has somewhat gained popularity among BDL researchers. Although most people associate “image classification” with supervised learning multiclass classification, it actually refers to a broader spectrum of machine learning applications. Rączkowski *et al.* [43] have proposed a BDL model named (ARA-CNN) which stands for Accurate, Reliable, and Active Bayesian Convolutional Neural Network for classification of histopathological images of colorectal cancer. Their model displays the level of uncertainty associated with each image, which can be used to detect mislabeled images. The proposed method, however, was only evaluated on one balanced dataset. Therefore, it will be interesting to see how it performs on other

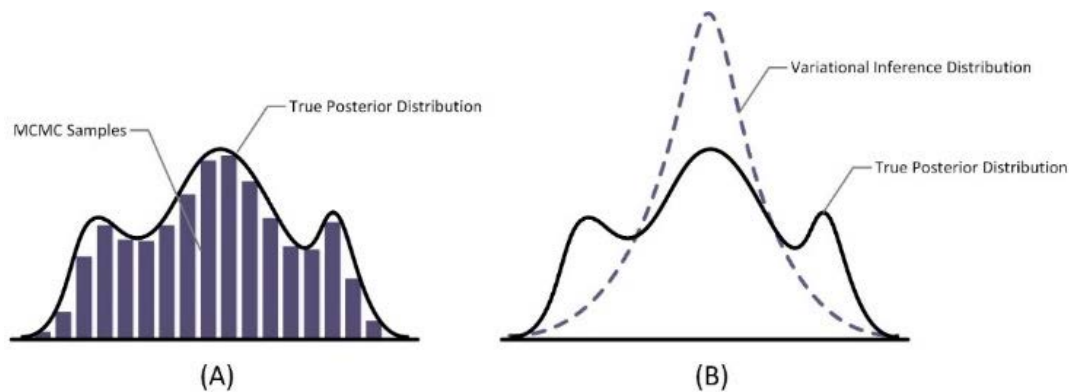


FIGURE 6. The difference between MCMC (A) and VI (B), inspired by [17].

datasets. Song *et al.* [44] used BDL to classify intraoral cancer images, and the uncertainty was used to propose more reliable readings from the images. The proposed model was based on VGG19 and used the MC-Dropout method as a Bayesian approximation. Their study had limitations in that it only applied the proposed method on a single dataset of 2350 cheek mucosa images, and there were few performance details compared to other published studies. Filos *et al.* [45] tested different Bayesian methods such as Mean-Field variational inference (MFVI) and MC-Dropout to investigate the robustness of BDL models for diabetic retinopathy classification. The VGG architecture was used as a basis for most of the tested models. According to the results, for the retinopathy dataset, MC-Dropout and ensembles methods outperformed MFVI, and combining the two methods can improve performance. The strength of the paper comes from a comparison of multiple different models and their performance although only a single medical dataset was used with some other non-medical datasets. Yadav *et al.* [46] used convolutional BDL to classify Parkinson's disease using functional magnetic resonance imaging (fMRI). They designed a model of BDL to feed slices of fMRI images to the network, and the proposed network was comparable to LaNet-5, but for 3D fMRI images. The weak point of this work is the author only proposed a single model, and details of the Bayesian inference implantation were not given, although details of the test implementations and hardware were given. Liu *et al.* [47] used BDL for UQ in a chest X-Ray image classification task using Stochastic Weight Averaging Gaussian (SWAG) [48]. They used different CNN-based architecture, including DenseNet, ResNet, ResNeXt, and SENet, each with a different number of layers, and they used the CheXpert dataset that contains over 200 thousand X-ray images. The paper's strength is that it used multiple models to compare performance, although it did not publish results for all lesions in the CheXpert dataset that was used for testing. In another study, Khan *et al.* [49] adopted BDL to predict breast cancer existence and achieved significant results for sensitivity. Multiple tasks, such as classification, segmentation, and image

enhancement have been implemented in this work, but no details of implementation have been provided. By the same token, Kwon *et al.* [50] investigated both aleatoric and epistemic uncertainties while employing BDL for UQ for image classification using decomposition moment-based prediction. They used three segmentation datasets, two of which are Ischemic stroke lesion segmentation and one retinal image dataset to extract vessels. The proposed method used variational inference to sample from the posterior distribution. It is worth mentioning that that work was the continuation of a previously published work, by the same authors, two years earlier. In another related medical image classification study, Khairnar *et al.* [51] used Convolutional BDL on breast histopathological images for uncertainty quantification in classification. In their proposed approach, they used activation functions with learnable parameters to address CNN's shortcoming in both the Bayes and non-Bayesian frameworks. This work differs from most others in that it used variational inference method in convolutional layers too rather than fully connected layers alone. In another study by Van Molle *et al.* [52], DBL was used for quantifying uncertainty in the classification of skin lesions. The proposed method used ResNet50 as a feature extraction method and MC-dropout was used as a Bayesian approximation to investigate uncertainty associated with images. The authors have estimated uncertainty per class for the proposed method, but the shortcoming of this work is the performance comparison with some existing work. Oloyede *et al.* [53] used BDL for Covid-19 X-ray images classification. The proposed BDL models were compared to a non-Bayesian model to investigate the applicability of Bayesian CNN with comparison to classical CNN. The paper used variational inference to sample from the posterior distribution. The limitation of this paper was that it only used a single very small dataset of 50 images, and results were not compared to existing methods. In another work, Gour *et al.* [54] deployed BDL for uncertainty-aware of chest X-ray images classification for Covid-19. The presented method was based on EfficientNet and used MC-Dropout as a Bayesian approximation

TABLE 1. A summary of published articles that used BDL for medical image classification.

Reference no.	Year	Bayesian technique	Purpose	DL model/ architecture
[43]	2019	Variational dropout	classification of histopathological images of colorectal cancer	CNN with inspiration from ResNet, DarkNet
[44]	2021	MC-dropout	Classification of intraoral cancer images	VGG19
[45]	2019	MFVI, MC-dropout	classification of diabetic retinopathy	VGG
[46]	2021	MC-dropout	Classification of Parkinson's disease using fMRI	CNN similar to LaNet-5 (For 3D images)
[47]	2020	MC-dropout	classification of chest X-Ray images	DenseNet201, ResNet152, SENet154, ResNet, ResNeXt
[49]	2017	Variational inference	Breast Cancer prediction	Own implementation of CNN
[50]	2019	Variational inference	Uncertainty Quantification in image classification	Own implementation of CNN
[51]	2020	Variational inference	Classification and UQ of Breast Histopathology Image	Own implementation of Bayesian CNN
[52]	2019	MC-dropout	classification of skin diseases	ResNet50
[53]	2021	Variational Inference	Classification of Covid-19 from chest X-ray images	Own implementation of CNN
[54]	2020	MC-dropout	Uncertainty aware classification of Covid-19 from chest X-ray images	EfficientNet

method for estimating posterior. The results were compared to multiple existing methods which outperformed them. Table 1 summarizes the above-mentioned studies that applied BDL methods for medical image classification with their Bayesian technique, purpose, and model structure used.

2) IMAGE SEGMENTATION

Another image processing task that is particularly useful in medical imaging is image segmentation. Image segmentation tends to partition an image into segment(s) that share similar properties, resulting in a simpler form of the image that can be further processed and analyzed. Medical image segmentation is used to find regions of interest such as finding objects, volume, shape, and boundaries of objects. Objects can be different organs, tissues, bones, external items and they

can have different shapes and sizes. This task is particularly useful for radiologists since they deal with various kinds of images, and it can be used as a computer-aided technology as a second opinion by radiologists. Several papers have been published in recent years using different architectures of BDL for medical image segmentation. U-Net is the most popular deep learning architecture for the image segmentation task, and it is a fully CNN network. It mainly consists of constrictive and expansive parts. The constrictive part down-samples an input, whereas the expansive part over-samples an input. Figure 7 illustrates different types of image segmentation.

Orlando *et al.* [55] used BDL for photoreceptor layer of cells segmentation from Optical Coherence Tomography (OCT) images. They proposed a model for segmentation based on the U-NET architecture. It used the Bayes model to estimate epistemic uncertainty and error rate for areas of interest. The proposed model used MC-dropout as a Bayesian approximation technique to define epidemic uncertainty. Implementation details of the proposed method are specified in the paper. The only shortcoming of the proposed work is the use of a single small dataset. In another work, Roy *et al.* [56] adapted BDL in QuickNAT architecture for the entire brain segmentation structure-wise quality control using MRI T1 images of the brain. The QuickNat has a U-shaped architecture with two-dimensional fully-CNN that segments slices of an image using both coronal and sagittal axes. The proposed method also used MC-Dropout for sampling from the posterior distribution as a Bayesian approximation. The MC-dropout is used to define the proposed model's voxel-wise (volume pixel of 3D images) uncertainty. What makes this work stand out from the crowd is that the proposed model uses the whole brain instead of a single region of the brain. McClure *et al.* [57] used BDL in brain segmentation using MRI images. In addition to the MC-dropout, the proposed model adapted spike-and-slab dropout to acquire dropout probability and individual uncertainty associated with weights. The model predicted uncertainty for voxel-wise error rate to predict the quality control manual annotation. The strength of the paper comes from the implementation of different Bayesian methods and the performance comparison of methods. Hiasa *et al.* [58] used Bayesian-CNN based U-Net for muscle segmentation from Musculoskeletal for CT scan images. The suggested model used MC-Dropout as a Bayesian approximation for segmentation as well as quantifying uncertainties in the model. The model performed better than previous work on two datasets used by authors. Furthermore, the proposed technique investigated uncertainty in multi-class for organ segmentation and how it may be utilized to execute predicted segmentation without the requirement for ground truth data, as well as reducing manual annotation for samples in the active learning task. Ma *et al.* [59] used BDL in a dense U-Net for segmentation of pancreas using the statistical shaped method. The Bayesian VI approach was used in the proposed model to overcome statistical shape model shortcomings such as localization and initial condition sensitivity. According to the authors, the results outperformed

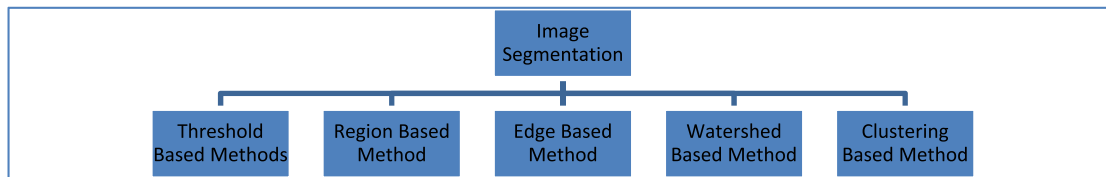


FIGURE 7. Image segmentation types.

the state-of-the-art results from previously published works. The paper's flaw is that it does not describe how the proposed model and hyperparameters are implemented.

In another work, Saidu *et al.* [60] proposed an active learning method based on BDL U-Net architecture for image segmentation. The proposed method used the CNN model for images with MC-Dropout as Bayesian approximation to predict model uncertainty. The proposed model was applied to four different medical image datasets. The use of active learning for segmentation tasks and the application of the model to different medical datasets are two of the work's strengths. Sedai *et al.* [61] used BDL for UQ for semi-supervised segmentation tasks of OCT images of retinal cells. In semi-supervised learning, only a limited amount of data are labeled and the rest of the data are unlabeled where the model is trained on them to make predictions. The proposed model was trained on labeled data for segmentation using BDL, which produces soft segmentation labels with their uncertainty that are then applied to unlabeled data. The proposed method was based on dense U-Net architecture and used MC-dropout as a Bayesian approximation method. The strength point of the proposed model is the deployment of BDL for semi-supervised tasks. Sander *et al.* [62] used BDL for automatic segmentation using dilated CNNs to produce a mask for segmentation and maps of spatial uncertainty. The generated spatial uncertainty map was used on automatic segmentation which increased the performance of segmentation while having assistance from the human intervention for high uncertainty regions. The proposed method used MC-dropout as a Bayesian approximation to generate spatial uncertainty maps. On the bright side, the paper well-described hyperparameters in detail, but it did not compare the method to other state-of-the-art methods. Sedai *et al.* [63] used BDL for segmentation and uncertainty estimation of OCT images for retinal layers. The Bayesian method was used to quantify uncertainty for pixel-wise OCT images for the segmentation task. Because high uncertainty is nearly always the result of erroneous segmentation, the BDL uncertainty maps come in helpful for finding wrongly segmented pixels and areas of images. The proposed method used MC-dropout as a Bayesian approximation method to generate the uncertainty map for segmentation. The strong point about this paper is that the authors have mentioned details of hyperparameters and compared their proposed method performance with two existing methods. Jena *et al.* [64] used BDL for diseases segmentation and uncertainty associated with it. The authors used three medical

imaging datasets for the brain, cells, and chest diseases. The authors have clearly defined hyperparameters in detail and used several datasets to examine the performance of the proposed method, however, it has flaws in that it does not compare their model's performance to existing state-of-the-art methods. Antico *et al.* [65] used BDL in the segmentation of knee arthroscopy for ultrasound images. The proposed model was based on Bayesian CNN and used MC-dropout as a Bayesian approximation method. Besides using BDL for segmentation, the model took advantage of MC-dropout to find pixel-wise uncertainty of the image. The model was tested on the femoral cartilage of knee image for ultrasound and MRI, which outperformed classical CNN. The performance details of the proposed methods were not compared to other state-of-the-art methods, although different datasets and hyperparameters were used. Liu *et al.* [66] used BDL for segmentation of the Amygdala Subnuclei region of the brain. The proposed model used Bayesian CNN for 3D images and deployed MC-dropout as Bayesian approximation for segmentation and uncertainty quantification. Because targeting and segmenting sub-regions with high accuracy is a hard task for classical deep learning, BDL was deployed for such tasks. It achieved better results compared with the state-of-the-art methods, especially with the presence of an uncertainty quantification map. Liu *et al.* [67] used BDL for automatic segmentation and uncertainty estimation for the prostate zonal of peripheral zone and transition zone. The proposed model used Bayesian CNN with attention mechanism and MC-dropout as a Bayesian approximation method for finding zonal segmentation uncertainty. It is worth mentioning that this is the only method that used the attention mechanism for segmentation. Garifullin *et al.* [68] used BDL to segment images of diabetic retinopathy lesions. The proposed method used Bayesian CNN with MC-dropout as a Bayesian approximation method to estimate pixel-wise uncertainty of segmentation of four types of lesions. The work's merits include good visualization charts that can rapidly grab the reader's attention. However, there is not much in terms of model performance when compared to other methods presented. Table 2 summarises the above-mentioned studies that applied BDL methods for medical image segmentation with their Bayesian technique and study purpose. Largent *et al.* [69] utilized BDL for brain segmentation of preterm infants depending on post-hemorrhagic hydrocephalus. The study used T2 MRI brain images which refer to MRI images that highlight fat and water content in the body. The study adopted MC-Dropout method

TABLE 2. A summary of articles that used BDL for medical image segmentation.

Reference no.	Year	Bayesian technique	Purpose
[55]	2019	MC-dropout	Segmentation of the photoreceptor layer in pathological OCT scans
[56]	2019	MC-dropout	Segmentation of entire brain using MRI T1 images
[57]	2019	MC-dropout	Segmentation and UQ of brain MRI images
[58]	2019	MC-dropout	Muscle segmentation using CT scan
[59]	2018	Variational inference	Segmentation of pancreas using CT scan
[60]	2021	MC-dropout	Active learning image segmentation and UQ
[61]	2019	MC-dropout	Semi-supervised segmentation and UQ of OCT images
[62]	2019	MC-dropout	Segmentation and UQ for cardiac MRI
[63]	2018	MC-dropout	Segmentation and UQ for OCT's retinal
[64]	2019	Not known	Segmentation and UQ for chest diseases
[65]	2020	MC-dropout	Segmentation of knee arthroscopy
[66]	2019	MC-dropout	Segmentation and UQ of the Amygdala Subnuclei region of the brain
[67]	2020	MC-dropout	Segmentation and UQ of the prostate
[68]	2021	MC-dropout	Segmentation and UQ of diabetic retinopathy lesions
[69]	2022	MC-dropout	Segmentation of brain for preterm infants

of Bayesian approximation to sample from the posterior distribution. The proposed method was tested on 27 subjects that were manually labeled for different parts. In addition to the segmentation results, the proposed Bayesian method was used to generate uncertainty maps associated with output. Several results were demonstrated in the presented work for various hyperparameters and models, indicating the difference between various configurations and which configuration performs better. However, the proposed method was not compared to other state-of-the-art methods and they used a small dataset for training and testing.

3) IMAGE REGISTRATION

Image registration is the process of transforming two or more images into similar coordinates that are geometrically aligned to reduce the differences between them. This process is essential for analyzing and preprocessing images that are generated from different sources, under different conditions, and at different times. This process is particularly useful in medical image analysis. This is due to the fact that the different conditions of images that deep models can be trained on and the ability to test generated everyday images by a different radiologist using different angles, lightning, and

TABLE 3. A summary of articles that used BDL for medical image registration.

Reference no.	Year	Bayesian technique	Purpose
[70]	2016	Variational inference, MCMC	UQ in image registration
[71]	2019	Variational inference	Image registration for deformable medical images
[72]	2020	MCMC	Image registration for unsupervised deformable brain MRI
[73]	2021	MC-dropout	Brain image registration

other factors. In this regard, Le Folgoc *et al.* [70] applied BDL to quantify uncertainty in medical image registration using a sparse Bayesian model. The presented model used VI as the approximate Bayesian method for sampling from the posterior distribution and used MCMC as the exact “asymptotically” Bayesian sampling method. According to the authors, the VI mechanism does well in the inference process similar to MCMC, however, the uncertainty estimation of VI is not as good as MCMC. The strength of their paper comes from using both VI and MCMC, which is rarely used by the research community.

Deshpande *et al.* [71] employed BDL for deformable medical image registration. Deformable images registration is an essential task in medical imaging that has a variety of applications, including multi-modality image fusion, temporal changes in structure, and so on. The proposed work used BDL for corrupted images by nonlinear geometric distortion. Because of Bayesian model characteristics, the model parameters were tuned to have a probability distribution, which increases the performance of the image registration. The proposed methods were investigated on a few datasets of deformed images. In another similar application, Khawaled *et al.* [72] used BDL for unsupervised deformable image registration of brain MRI. Unsupervised DL models for deformation image registration are trained on available data to estimate and measure the deformation through calculating similarity and differences target and other images. The trained deep learning model was then used on other data that was apart from training. The posterior distribution of the model evades overfitting even when the size of the dataset is small. The proposed model used stochastic gradient Langevin dynamics to sample from posterior as a Bayesian method for image registration, and uncertainty associated with deformed images were quantified. The paper is one of the few works that have deployed MCMC for Bayesian sampling. Cui *et al.* [73] used Bayesian CNN for brain image registration. The suggested method used MC-dropout as a Bayesian method to sample from posterior for the registration task as well as producing the geometric uncertainty map for the uncertainty associated with the registration process. Table 3 summarises the above-mentioned studies that applied BDL methods for medical image registration with their Bayesian technique and study purpose.

4) IMAGE RECONSTRUCTION AND ENHANCEMENT

Image reconstruction is the process of creating images from incomplete or scattered data which can either be two or three-dimensional images. The incompleteness or scattered data can be generated or caused by different sources such as radiation reading for medical images or noisy object removal in an image. It is particularly useful in medical applications for generating useful images, which in some cases may require applying some mathematical methods. For three-dimensional images such as CT and MRI, this technique has the ability to create three-dimensional images as brain images from a variety of two-dimensional images. This technique can also be used to sharpen images or sharpen edges for objects in images as a preprocessing technique for other image processing tasks like image segmentation. On the other hand, image enhancement is the process of modifying images to result in a visually better image or improving them for additional processing and analysis. Image enhancement results in an accurate visual representation of images and improves the quality of image features for image processing tasks. This task is achievable through different mechanisms such as image sharpening, noise removal, and adjusting image intensity.

Schlemper *et al.* [74] used BDL to quantify the uncertainty associated with the model for reconstructing MRI images. Two datasets of cardiac MRI were used in this work. The uncertainty in the proposed model was particularly effective in high-uncertainty regions, in which the model might fail to construct the image from available data. The authors used the MC-dropout method as a Bayesian approximation to determine the epistemic uncertainty in the proposed model. The authors also investigated the relationship between the uncertainty and error associated with predicted images and found a correlation between the two, indicating that the error in the model mostly comes from high uncertainty regions in images. The strength point of the paper is that several different models and parameters were implemented with visualization of their performances. These models, however, were not compared to other existing states of the models. Du *et al.* [75] used BDL for multi-view visual image construction from human brain activities using functional-MRI (fMRI). The authors relied on visual stimuli and educed fMRI to represent the correlation between the two viewpoints. The published work used a linear model of BDL to identify the voxel correlation in addition to handling noise in data to avoid the overfitting of the proposed model. Marinescu *et al.* [76] used Bayesian approach generative models for image reconstruction. The purpose of using the generative model in this work was to overcome the distribution shift of test data. This is performed by having a single image generator for different image reconstruction tasks. The proposed model was tested on three datasets among which were two medical datasets: chest X-ray and brain MRI datasets. The proposed method used variational inference for sampling from the posterior distribution. The performance of different hyperparameters for models was tested and compared to other state-of-the-art

models. Yang *et al.* [77] used the Bayesian method for a Positron Emission Tomography (PET) scan reconstruction. They proposed a Multilayer Perceptron (MLP) patch-based model for PET scan reconstruction using VI for model regularization. This method was used to increase the readability region of the radioactive tracer that was added to the human body's liquids. The results of the paper were well-visualized, and details of model hyperparameters were described, however, not many details were mentioned compared to other existing methods. Barbano *et al.* [78] used a Bayesian model for knowledge transfer of learning outcomes of iterative reconstruction. The authors used VI for Bayesian approximation to sample from the posterior distribution. The proposed method firstly trained the model for reconstruction on supervised data with ground truth labels using Bayesian VI and then fine-tuned the model parameters based on samples adaptation measurements using unsupervised data. The model also provided uncertainty quantification for the unsupervised data. Tanno *et al.* [79] used Bayesian techniques for image enhancement to identify different parts of uncertainty. To quantify the uncertainty associated with model parameters and data noise "aleatoric and epistemic," they applied their proposed method to diffused MRI images for super-resolution tasks. The outputs were predicted based on these two uncertainties. As in other publications, the authors concluded that uncertainty quantification enhances the performance of models, particularly for shifted distribution cases. In addition, the area with a high error rate mostly has high uncertainty which indicates the positive correlation between uncertainty and error. In different work published earlier by Tanno *et al.* [80] which the work [79] was based on, they used Bayesian methods for diffused MRI super-resolution to quantify uncertainties that are a result of noise in data and model parameters. Li *et al.* [81] used BDL for time-series data for high resolution of fluorescent images with high density. The proposed method was used for structure reconstruction of super-resolution fluorescence microscopy. The authors used experimentally calibrated parameters to overcome the problem of overfitting in the model. The suggested model used MC-dropout as a Bayesian approximation method to sample from the posterior distribution. Table 4 summarizes papers that used BDL methods for medical image reconstruction and enhancement with their Bayesian technique and research goals.

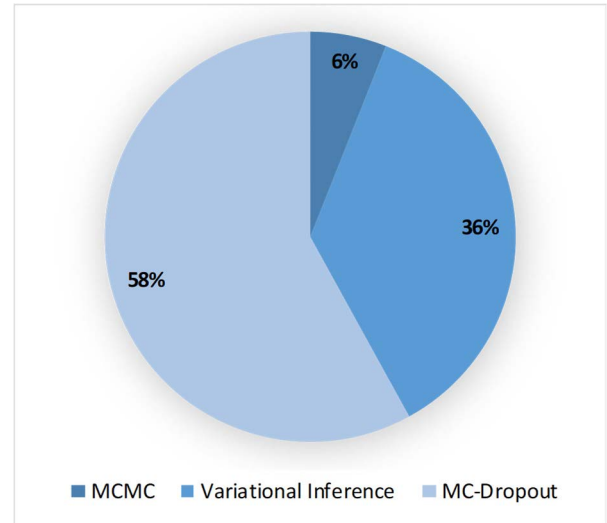
5) OTHER TASKS IN MEDICAL IMAGING

In addition to the above-mentioned task in medical imaging, there is also some published work that covers different tasks for medical imaging. Hassan *et al.* [82] used incremental learning for cross-domain adaptation for retinopathy to extract anomalous retinal in optical coherence tomography (OCT) images. The authors used only few-shot learning instead of training for a long period of time. The benefit of using incremental is that it requires less training for new images without the need for models training for a long period and with no need for past training images. They used a

TABLE 4. A summary of articles that used BDL for medical image reconstruction and enhancement.

Reference no.	Year	Bayesian technique	Purpose
[74]	2018	MC-dropout	UQ for MRI image reconstruction
[75]	2018	Variational inference	Multi-view image construction from human brain activities using fMRI
[76]	2020	Variational inference	Generative image reconstruction
[77]	2018	Variational inference	PET scan reconstruction
[78]	2021	Variational inference	Knowledge transfer of learning outcomes of iterative image reconstruction
[79]	2021	Variational dropout	UQ for neuro image enhancement using dMRI
[80]	2017	Variational dropout	UQ for noise associated with input using dMRI
[81]	2018	MC-dropout	time-series data for high-resolution fluorescent images

Bayesian multi-object function to enable the network to know the gap and the semantic relation of new and trained images. Multiple well-known CNN architectures were used in their work such as MobileNet, ResNet-50, ResNet-101, and VGG-16. Gal *et al.* [83] worked on active learning for image data using BDL models which was the first attempt to apply BDL in active learning. They used Bayesian CNN for their proposed work. Despite the fact that the paper was not entirely focused on healthcare, they did use a skin cancer (melanoma) image dataset (ISIC2016). They also fine-tuned the VGG-16 which was pre-trained on ImageNet. Mahapatra *et al.* [84] used BDL for active learning tasks in both classification and segmentation tasks. They used Generative Adversarial Networks (GANs) to produce chest X-Ray images that have chest lesions features. The proposed model used the most useful samples for training to have more informative features. BDL was used to find and use samples generated by GAN with the most informative features and characteristics of lesions. They took advantage of using VGG16 and ResNet18 which were pre-trained on ImageNet. The paper did not mention the sampling method for BDL, although it described other hyperparameters of the model. Gou *et al.* [85] used BDL for classification and segmentation of Subarachnoid hemorrhage disease using CT scan images of the head. The purpose of the BDL in the proposed method was to estimate the uncertainty associated with model predictions and based on that estimation, professionals can decide the reliability of the predicted output. Moreover, the authors implemented a model for semi-supervised learning to illustrate the regions with uncertainty on the images for segmentation. The paper used MC-Dropout as a Bayesian approximation for sampling from posterior distributions and used EfficientNet structure for building a classification model. The models' performance for classification and segmentation was not compared against other state-of-the-art methods, although multiple different hyperparameters

**FIGURE 8. Ratio of Different BDL techniques used for medical image processing.**

were tested. Nakao *et al.* [86] employed BDL for anomaly detection in chest PET/CT scans using F-fluorodeoxyglucose tracers. The proposed model used MC-Dropout as a Bayesian approximation method to sample from the posterior distribution. The proposed method was evaluated on only 34 images from a small dataset of less than 1900 PET/CT scans with no abnormalities. In terms of the ROC curve, the proposed model performed very well and reached 99.2%, but not so well in other measurements. The obtained results were compared to a limited number of relevant works, although other works did not use the same number of images, which indicates why there is a performance disparity.

BDL has been used in a variety of medical imaging activities; however, not all of them have received equal attention. As it can be seen in section 6.1 that most of the published work was for image classification and segmentation tasks. More investigations are needed in deploying BDL for other medical image tasks, particularly for uncertainty quantification, interpretation, and trustworthiness tasks. Another important point to be highlighted is the use of Bayesian methods for medical imaging tasks, as illustrated in Figure 8 the majority of the reviewed works have used MC-Dropout. In general, images have a very large number of features for consideration which makes them complex structures, and MC-Dropout is the fastest method to converge, so most researchers use it as their primary method. Only a few papers have used the MCMC method since it takes a lot of computational power for the model to reach an equilibrium state, although it has more accurate results compared to other methods.

B. MEDICAL SIGNAL PROCESSING

Medical signals are the result of a living being's subdermal tissues' accumulated action potentials. Medical signals are records in space, time, or space-time of a biological

occurrence in a living being, such as heart beating, contracting muscle, and brain electrical activities. The electrical, chemical and mechanical activity that occurs throughout this biological process can provide signals that can be detected and evaluated. As a result, medical signals include useful information that can be used to analyze the physiological mechanisms underlying a specific biological event or system, as well as for medical diagnosis. These signals are temporally based actions that can be measured either electrically, as in Electroencephalography (EEG), which records electrograms from brain electrical activities; or magnetically, as in Magnetoencephalography (MEG), which records brain activations using magnetic fields. The nature of these signals is random, and they cannot be accurately predicted. These signals are crucial to determine how does brain or heart is functioning which mostly cannot be evaluated through medical imaging. Medical signal processing has gained popularity among BDL researchers. In this regard, Chai *et al.* [87] used BDL to classify mental fatigue using EEG signals. The authors used Principal Component Analysis (PCA) as a preprocessing technique for dimensionality reduction to reduce the amount of EEG channels from 26 to only 6 and used power spectral density for feature extraction. The experiment used eyes open and eyes closed cases to measure the fatigue status of the brain. The paper did not go into details regarding how the BDL was used, such as which sampling method was used to sample from the posterior distribution, but instead focused on the PCA in more detail. However, Chai *et al.* in earlier work [88] used BDL as a classifier to compare EEG feature extraction for fatigue detection. They used Power Spectral Density (PSD) for feature extraction, which has been used by many other researchers. For comparison, they used different feature extraction techniques including Power Spectral Entropy (PSE), Wavelet, and Autoregressive (AR). The authors concluded that AR has better performance than other techniques used for feature extraction. Interestingly, the authors have not indicated whether or not the AR will perform better when using classical DL. The same group of researchers Ngo *et al.* [89], published another work using BDL for the classification task of Nocturnal Hypoglycemia based on EEG signals. Hypoglycemia is a common condition that diabetes patients face during sleeping at night. The proposed method did not specify the sampling method or distribution type for the BDL. In another study, Handojoseno *et al.* [90] used BDL to predict the occurrence of freezing of gait before happening in Parkinson's patients using EEG signals. Freezing of gait is a disorder on gait that affects Parkinson's disease patients when they lose the ability to move unexpectedly and momentarily, resulting in falls even when patients have the intention to move and walk. In their proposed method, the authors used two preprocessing techniques: directed transfer function for brain connectivity actions and Independent Component Analysis (ICA) to separate the components of a signal. The BDL was used to predict the occurrence of freezing the gait with having five seconds before its occurrence. The Bayesian method was

used as a regularization technique to prevent the model from overfitting. The author suggested that additional investigation with more data and applying more methods in real life is needed, in order to prevent freezing of gait occurrences for Parkinson's patients. Although the authors included implementation details on the number of layers and neurons used, they did not describe how the BDL was adapted in the proposed method, such as which sampling method was used. Fruehwirt *et al.* [91] used BDL to detect Alzheimer's disease severity using neurophysiological markers EEGs. BDL was used to create a multivariate predictor to detect the severity of the disease using EEG markers. In addition, to predict the output distribution, BDL was used in the proposed model to quantify uncertainty. Two sampling methods were used in the proposed models, namely MC-dropout as Bayesian approximation and Hamiltonian Monte Carlo (HMC) as an exact sampling technique to sample from the posterior distribution. Results from the different models showed that BDL with HMC sampling outperformed MC-dropout and classical DL. The reason for this is that the distribution is approximated for MC-dropout, which underestimates the predictive uncertainty, according to the authors. The strength of the proposed method comes from using HMC. The paper, however, did not compare the gained results with other existing methods. Lee *et al.* [92] adopted BDL for artifact removal from EEG signals. ICA was used to extract independent components from EEG signals and other information that was not part of independent components were treated as artifacts and removed. BDL was used for classification purposes alongside ICA. The MC-dropout was used as a Bayesian approximation to sample from the posterior distribution. In addition, the proposed method used the attention mechanism to improve the performance by concentrating on a region of interest. However, the authors did not compare the archived performance form datasets with state-of-the-art models.

In addition to EEG, several works have been done on electrocardiogram (ECG) signals. ECG is used to test the activities of the heart through an electrical signal that is produced by heartbeats and recorded by sensors attached to the skin. For this purpose, Aseeri *et al.* [93] used BDL to classify the arrhythmias "abnormal rhythm of heart beats" through ECG signals. The proposed method estimates the uncertainty associated with the predicted output. The proposed method used MC-dropout as a Bayesian approximation method to quantify model uncertainty. In addition, the Gated Recurrent Units (GRU) method was used instead of the vanilla recurrent neural network, resulting in improving the performance of the proposed method. The proposed method performed better than other state-of-the-art models as described in related work sections of the paper. In another work, Hua *et al.* [94] used BDL for the interpretability task for ECG data since the classical DL cannot measure its uncertainty to determine which and how each input affects the predicted output. Since decision-making in the medical field can be a very sensitive task, machine learning and DL models should have ways describing how output(s) are

predicted from input. For this purpose, the authors extracted features from the ECG signals and fed extracted features to the BDL model to describe how each feature affected the output. The VI was used as a Bayesian approximation to sample from the posterior distribution, and different types of priors such as a horseshoe, half-Cauchy, and Gaussian were investigated to perform the evaluation and comparison of features for a classification task. The paper of Hua *et al.* [94] is one of only a few papers that used different types of prior distributions in BDL. Jagannath *et al.* [95] also used BDL to extract the fetal cardiac signals. The fetal ECG can be detected in an advanced stage of fetal growth. In a clinical test, it is hard to detect anomalies in heart functioning because of the lack of advanced tools for detecting signals to identify anomalies. The authors used BDL to remove maternal ECG from fetal ECG to improve the quality of a signal. However, the details of BDL hyperparameters were not mentioned. Das *et al.* [96] used BDL to detect atrial fibrillation from Photo plethysmography. The proposed method was used to detect atrial fibrillation from noisy photoplethysmography signals using sensors from smart devices, such as smartwatches, to quantify the uncertainty associated with the predicted output. The proposed model used variational inference as an approximation Bayes method. The authors described model hyperparameters in detail along with the hardware used to run the experiments; and compared results with some recently published works. In another similar study, Belen *et al.* [97] used Bayesian methods to quantify uncertainty for atrial fibrillation classification using ECG signals. The proposed method used VI as a Bayesian approximation to sample from the posterior distribution. The authors described the structure and hyperparameters of the models in charts and compared the result with some of the existing methods. Their proposed method; however, could not achieve the state of the performance according to published results. Table 5 summarizes papers that used BDL methods for medical signal processing with their Bayesian technique and research goals.

C. MEDICAL NATURAL LANGUAGE PROCESSING (NLP)

NLP is a type of machine learning, which refers to the techniques for analyzing speech and natural language. The NLP does use Recurrent Neural Networks (RNN) to process information since this information has a sequential shape. NLP aims to enable the computer to understand the content of the text and the structure of the language to extract useful knowledge and information from a given document. When NLP is used in the healthcare field, it can help predict patient outcomes, enhance hospital triage systems, and produce diagnostic models that detect early-stage chronic disease. Physicians and medical staff are extensively adopting medical reports of patients to keep track of patients' status to diagnose diseases. Proper use of these documents is an essential task since it has all information about the patient's status. Several types of research on medical NLP have been done using classical deep learning techniques. Some examples of medical NLP using classical deep learning are: predicting the

TABLE 5. A summary of articles that used BDL for medical signal processing.

Reference no.	Year	Bayesian technique	Purpose
[87]	2016	Not known	Classification of mental fatigue using EEG signals
[90]	2018	Not known	Predicting of freezing of gait occurrences among Parkinson's patients
[91]	2018	MC-dropout, MCMC (HMC)	Detection of Alzheimer's disease severity using neurophysiological markers EEG
[88]	2015	Not known	Classification of EEG signals
[89]	2019	Not known	Classification of Nocturnal Hypoglycemia based on EEG
[92]	2020	MC-dropout	Artifact removal from EEG signal
[93]	2021	MC-dropout	Classification of the arrhythmias using ECG signal
[94]	2021	Variational inference	Interpretability task for ECG data
[95]	2019	Not known	Extraction of the fetal cardiac signal
[96]	2020	Variational inference	Detect atrial fibrillation from Photo plethysmography
[97]	2020	Variational inference	UQ for atrial fibrillation classification using ECG signals

length of stay in hospital for admitted patients [98], extracting medication and drugs from patients' reports [99], classification of radiological reports [100], and lesion area detection using CT reports [101]. Despite this, we are not aware of any published work on medical NLP that has used BDL.

D. ELECTRONIC HEALTH RECORD (EHR)

EHR is an organized collection of patients' electrically stored data. These records are used to keep track of patients' status, which are shared between medical professionals within the same network. EHR can store a large amount of data regarding patients such as patients' personal and physical information, medical history, radiology images, treatments, medication/drugs, and allergies. Effective use of this information by medical professionals is crucial since it can save time and lives in some cases. Many researchers have been working on mining EHR data to extract knowledge using machine learning and classical deep learning methods. However, only a limited number of published works have been done on EHR using Bayesian deep learning. Dusenberry *et al.* [102] used BDL to analyze uncertainty associated with the model they proposed for EHR. For sampling from the posterior distribution, the authors used variational inference as a Bayesian approximation. They also employed ensembles of classical RNNs and Bayesian RNNs to determine the model uncertainty, and they concluded that metrics used to measure the model performance could not identify the uncertainty associated with the model. The authors concluded that the Bayesian method has the capability of capturing uncertainty associated with models better than ensembles of models. The paper also indicated that Bayesian models can provide

TABLE 6. A summary of articles that used BDL for EHR processing.

Reference no.	Year	Bayesian technique	Purpose
[102]	2020	Variational inference	UQ for EHR
[103]	2019	MC-dropout	UQ for EHR
[104]	2020	Variational inference	Prediction of the admitted patients' outcome in temporal bases
[105]	2021	Variational inference	UQ for EHR
[106]	2021	Variational inference	Individual-based prediction of EHR

uncertainty as per individual features where ensemble fails to do so. In another related topic, Qiu *et al.* [103] used BDL for uncertainty quantification of EHRs. The proposed method quantifies the uncertainty associated with data, which refers to as noise “aleatoric uncertainty”. The proposed work used MC-dropout as a Bayesian approximation for sampling purposes. The authors concluded that reducing the number of records to decrease the noise would improve model performance. However, results obtained from the proposed model were not compared to other published papers. Deasy *et al.* [104] used the Bayesian approach with RNNs to temporally predict the outcome of admitted patients. The authors used variational inference as the Bayesian approximation method to sample from the posterior distribution. The proposed method used BDL to estimate uncertainty in the predicted outcomes. When the model’s uncertainty is low or when there are more data to predict, the proposed method predicted the status of patients more frequently; and when there is a high level of uncertainty associated with output, it predicted the status of patients less frequently. Although multiple models were tested for their performance, the obtained results were not compared against other existing works. Li *et al.* [105] used BDL and Gaussian processes for uncertainty estimation in EHR. The proposed model used MC to sample from posterior distribution 30 times for an instance. Since both BDL and the Gaussian process can estimate uncertainty using different methods, the combination of them is beneficial to take the advantage of both methods to capture model uncertainty. The proposed method was tested on EHR of heart failure, depression, and diabetes to estimate uncertainty. Diaz *et al.* [106] also used a Bayesian method with a logical neural network to focus on individual-based prediction on EHR. The logical neural network has the characteristic of combining both classical neural networks for the learning process and symbolic logic for identifying knowledge and reasoning. To the best of our knowledge, this is the first and only paper adapting the logic neural networks. Table 6 summarizes papers that used BDL methods for Electronic Health Records.

E. CLINICAL AUDIO AND VIDEO PROCESSING

Audio and video processing refers to the process of analyzing the content of audio or video in order to extract

knowledge, such as object detection and tracking video or sound recognition and noise cancellation in audio. Video and audio have limited application in healthcare data analysis compared to images. Nonetheless, they are indispensable and have a vital role in healthcare applications such as ultrasound video, endoscopy, and temporal speech disorder. Only a few studies have used Bayesian deep learning techniques in this field. One of those is the work of Bodenstedt *et al.* [32] who used BDL for surgical workflow video analysis adapting the active learning approach. The authors used MC-dropout as a Bayesian approximation method for sampling and used RNNs and LSTM models to analyze data. The main purpose of this study was to determine what data points were to be labeled “annotated” in each step of the surgery by using tools detection and operation stage segmentation. This paper is one of the very few that used Bayesian LSTM. By the same token, Liu *et al.* [107] used the BDL model to extract visual features for speech recognition for people with physical speech disorders. A speech disorder is common among people with special needs, which in some extreme cases makes it impossible to understand speech. In the proposed method, features from visual representation and audio features were fused as a multimodal approach for predicting speech and estimating uncertainty associated with the predicted output. The proposed method used variational inference to sample from the posterior distribution which is approximated as Gaussian.

Other than medical imaging, which has been covered in sections 6.2 to 6.6 of this paper, BDL has been used for a variety of tasks in healthcare. While some tasks, such as medical signal processing, have gained some popularity, others, such as medical NLP, have, to our knowledge, received no attention at all. It’s unclear why some tasks, despite their importance in healthcare, have not received the attention they deserve. In contrast to medical imaging, the majority of work in non-imaging tasks have used variational inference, which can clearly be seen in Figure 9. It seems that the MCMC method has not gained popularity for non-medical imaging tasks, this is because of its computational complexity.

VII. BDL IN DISEASES DIAGNOSTICS/DETECTION

The process of determining which disease or condition is causing a patient’s symptoms and signs is known as medical diagnosis. For disease detection and diagnosis, a large number of researches in machine learning and deep learning have been developed. This section covers some diseases that have gotten more attention than others using deep learning techniques, especially BDL. It is worth noting that only a few published works for each disease or type of the disease are discussed in this section, and some diseases may have a larger number of papers that are not covered in this section. The following are some of the common diseases for which the BDL models were used.

A. COVID-19

Coronavirus disease (COVID-19), a contagious disease, is caused by the SARS-CoV-2 virus. The COVID-19

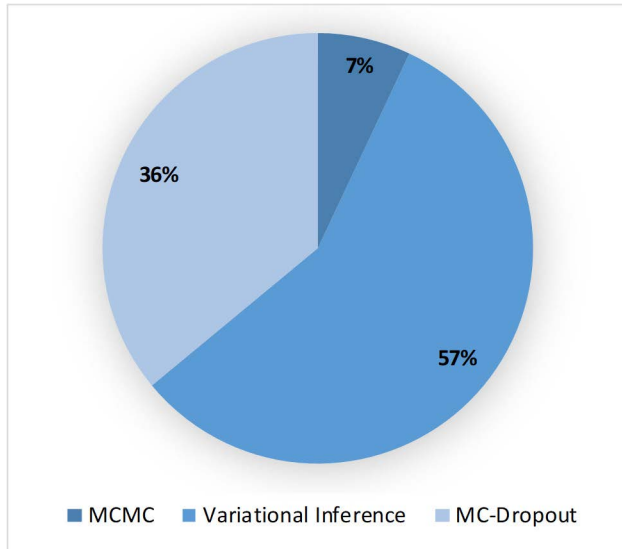


FIGURE 9. Ratio of different BDL techniques used for other medical tasks "non-image."

pandemic has impacted many individuals around the world in some form, and it has changed the way of life for most people. COVID-19 has become one of the world's biggest challenges, if not the biggest one, in the past two years. Thus, it got unprecedented attention that no other diseases in human history had seen. Numerous research studies have been published on COVID-19 in healthcare and other fields of study. Several papers have been published to detect and classify COVID-19 using machine learning and deep learning techniques. Yet, only a few papers have used BDL. Ghoshal *et al.* [108] used a Bayesian-based CNN model to estimate the uncertainty of chest X-ray images for COVID-19 detection. The authors adapted pre-trained ResNet50V2 as a baseline for their proposed model and fine-tuned its parameters with available data. They used the MC-Dropweight mechanism for sampling from the posterior distribution. Dropweight works the same way as dropout, but instead of dropping a neuron in a network "similar to setting all weights of inputs to zero", it drops only some of the neuron's weights with a given probability. Cabras *et al.* [109] also used BDL for estimation COVID-19 spreading in Spain for a period of 14-consecutive days. The proposed method was developed using the LSTM architecture and counts the "likelihood" of the available data using the Poisson distribution. Instead of using Gaussian priors, which the majority of BDL papers are based on, the proposed method uses Gamma priors to quantify model uncertainty. Ucar *et al.* [110] used SqueezeNet CNN architecture with Bayesian optimization to detect COVID-19 from chest X-ray images. Since the authors used a pre-trained SqueezeNet model, they fine-tuned model parameters with chest X-ray images. According to the authors, the model outperformed the state-of-the-art models used for COVID-19 detection. Oloyede *et al.* [53] deployed BDL to classify COVID-19 X-ray images and

compared the model with the non-Bayesian model. The purpose of this work was to investigate the applicability of Bayesian CNN with a comparison to classical CNN. The proposed BDL model achieved validation accuracy of 95%, while the classical CNN model reached 87% after one thousand epochs. In another work, Gour *et al.* [54] used BDL for uncertainty-aware COVID-19 classification of chest X-ray images. The suggested method was applied on three state-of-the-art datasets and achieved promising results. The results were compared with some other existing methods, which overachieved the existing methods. Loey *et al.* [111] employed CNN-based Bayesian optimization for the detection of COVID-19 using chest X-ray images. The purpose of Bayesian optimization in this paper was to fine-tune the hyperparameters of the network. The paper used balanced datasets for training that consisted of more than 10 thousand images of three classes, which are normal, COVID-19, and Pneumonia. Multiple scenarios of the proposed model were presented to compare results, and their best model performance reached 96%. In recent work, Niraula *et al.* [112] used BDL for modeling the spread of COVID-19 in 245 health zones in Spain using Spatio-temporal data. Because of the nature of the data, the suggested method used LSTM architecture and adapted Laplace approximation as VI to sample from the posterior distribution. The obtained results have not been compared to other state-of-the-art methods in the field due to the nature of the study and the data used.

B. CARDIOVASCULAR DISEASE

Cardiac disease refers to a disorder of heart functionality. Cardiovascular disease is a type of disease that affects the heart and blood veins. Smoking, high blood pressure, high cholesterol, an unhealthy diet, and obesity can all raise the risk of cardiovascular disease. Researchers have used machine learning and deep learning techniques to assist medical professionals in identifying and diagnosing cardiac diseases. However, there are only just a few published studies that have employed BDL for cardiac diseases. Sander *et al.* [62] employed BDL for segmentation of some parts of the human heart such as the left ventricle cavity, right ventricle, endocardium, and myocardium at end-diastole and end-systole. The designed method used MC-dropout for sampling, and it was assessed on MICCAI 2017 dataset, in which 100 cases were used; 75 during training and 25 for testing. Aseeri *et al.* [93] also used BDL to classify cardiac arrhythmias based on ECG signals. The suggested method used MC-dropout for sampling, and the model performance was tested on three datasets (MIT-BIH for 48 patients, St Petersburg INCART with 75 records for 34 patients, BIDMC dataset for 15 patients), which achieved an F1-score of 98.8%, 99.2%, and 97.2%, respectively. Jagannath *et al.* [95] deployed BDL to extract the fetal cardiac signals using ECG. The authors tested the proposed method on two datasets, namely Physionet and DaISy. Four different methodologies were used to test the performance. Their obtained

results, however, were not compared to other state-of-the-art methods.

C. CANCER

Cancer is a disease that arises when cells in the human body grow uncontrollably and unexpectedly in a certain region, with the potential to spread to other parts of the human body. Every living part of the human body is susceptible to cancer because it can affect any type of human. Human cells can get cancer, which can be inherited from parents, through exposing cells to prolonged harmful radiation, or from natural causes. Most types of cancer can be fatal when discovered in their advanced stages. Thus, early detection of cancer is crucial, as it can save a substantial number of lives each year. Several works have employed machine learning and active learning techniques to detect and diagnose this disease, with a few of them relying on BDL. For example, Liu *et al.* [67] used BDL for uncertainty estimation and segmentation of prostate cancer using MRI slices. The proposed method uses an attention-based Bayesian U-Net structure for segmentation and uncertainty estimation. The data for model training and testing were collected from two public sources to reach a total of 351 MRI scans. In another study, Song *et al.* [44] employed BDL to classify intraoral cancer images with uncertainty quantification for reliability. The authors used the MC-dropout method for sampling and pre-trained VGG19 as a backbone, which was fine-tuned using intraoral data with a probability of 0.5 as a dropout rate. The proposed method was tested on a dataset that contains images of 2350 cheek mucosa and the model reached an accuracy of 90%. Billah *et al.* [113] also deployed BDL to classify cancer images and quantify uncertainty for blood cells (lymphocytes) cancer (lymphoblast). The proposed model used the MC-dropout method for sampling, and each image had 50 samples. The authors used the (ALL-IDB2) dataset that has 260 images of cancer cells (lymphocytes), and the model reached a 94% accuracy rate.

D. PARKINSON AND ALZHEIMER

The brain is the main control unit in the human body, and it can be affected by several diseases that may affect its functionality. Alzheimer's and Parkinson's are two of those diseases that affect brain function. Both diseases may share some symptoms, but the main difference between them is that Alzheimer's does affect memory and language capabilities, while Parkinson's affects the brain functionality of thinking and some physical problems. A large number of researchers have used machine learning and deep learning methods to apply to different tasks such as classification and segmentation, but only a few researchers have used BDL. Amongst them is the work of Yadav *et al.* [46], which used Bayesian CNN for Parkinson's disease fMRI image classification. The proposed method used MC-dropout for sampling, and the data of 30 subjects were used to assess the performance of the proposed method, which resulted in an average accuracy of 97.92%. In another work, Roy *et al.* [56] employed BDL

in the QuickNAT architecture for automatic segmentation of the whole brain for Alzheimer's disease. They also used MC-Dropout for BDL sampling. The proposed method was tested on four datasets (MALC-15, ADNI-29, CANDI-13, and IBSR-18), where only the first one was used for training and the number of instances for each dataset is written next to the dataset's name. Their results showed that the mean dice score for datasets was between 81% and 88%. Handojoseno *et al.* [90] used BDL to forecast freezing gait for 5 seconds ahead of its occurrence using EEG signals for Parkinson's patients. The purpose of using BDL in the proposed method was to avoid overfitting of the model. In total, 16 patients' data sets were used; 11 for training and 5 for testing, with an average age of 70 years old. The proposed model achieved a sensitivity score of 85.86% and a specificity of 80.25%.

E. DIABETES

Diabetes is a chronic disease that affects the level of sugar (glucose) in the blood. During the digestion process in the stomach, food is broken down into glucose, which then circulates throughout the human body via blood circulation. The pancreas is responsible for controlling blood sugar; when it rises, the pancreas releases insulin to balance it. The cause of diabetes is that the pancreas cannot keep up with releasing enough insulin to maintain a balance of sugar in the blood. Several researchers have used machine learning and deep learning techniques for different task-related to diabetes. For instance, Filos *et al.* [45] compared the performance of different BDL techniques for diabetic retinopathy classification. The proposed model's performance was evaluated using the Kaggle diabetic retinopathy dataset, which has five classes based on the severity of the disease. The Ensemble MC-dropout model performed best and achieved accuracy and AUC of 92.4% and 88.1%, respectively. Garifullin *et al.* [68] proposed a BDL model to segment images of diabetic retinopathy using MC-dropout to estimate the uncertainty associated with predicted output. The authors used the IDRiD dataset for diabetic retinopathy and achieved a score of between 97.7% and 99.7% for the ROC-AUC metric.

Table 7 summarizes the reviewed paper in section 7. It can be seen that the most popular Bayesian method mentioned in this section is MC-Dropout. The reason is that most of the papers discussed in this section were using different tasks in medical imaging, and as stated in section 6, the MC-Dropout is the most popular method for imaging tasks. Another important point to mention is that the number of applications of BDL in healthcare is increasing. This is due to the importance of BDL model features, especially its ability to handle different types of uncertainties, which is a shortcoming of classical deep learning models.

VIII. DISCUSSION

The majority of the published works on BDL in healthcare applications have been reviewed in this paper, and their Bayesian methods have been demonstrated in the preceding

TABLE 7. A summary of articles that used BDL for EHR processing.

Reference no.	Year	Disease	Bayesian technique	Purpose
[108]	2020	COVID-19	MC-dropweight	Estimate the uncertainty of chest X-ray images for COVID-19 detection
[109]	2021	COVID -19	Not known	estimation COVID-19 spreading in Spain
[110]	2020	COVID -19	Not known	detect COVID-19 from chest X-ray images
[112]	2022	COVID -19	Variational Inference	Determine the spread of COVID-19 in Spain
[53]	2021	COVID -19	Variational Inference	Classification of COVID -19 from chest X-ray images
[54]	2020	COVID -19	MC-dropout	Uncertainty aware classification of COVID-19 from chest X-ray images
[62]	2019	Cardiovascular Disease	MC-dropout	Segmentation and UQ for cardiac MRI
[93]	2021	Cardiovascular Disease	MC-dropout	Classification of the arrhythmias using ECG signal
[95]	2019	Cardiovascular Disease	Not known	Extraction of the fetal cardiac signal
[67]	2020	Cancer	MC-dropout	Segmentation and UQ of the prostate
[44]	2021	Cancer	MC-dropout	Classification of intraoral cancer images
[113]	2021	Cancer	MC-dropout	classify cancer images and quantify uncertainty for blood cells
[46]	2021	Parkinson	MC-dropout	Classification of Parkinson's disease using fMRI
[56]	2019	Alzheimer	MC-dropout	Segmentation of entire brain using MRI T1 images
[90]	2018	Parkinson	Not known	Predicting of freezing of gait occurrences among Parkinson's patients
[45]	2019	Diabetes	MFVI, MC-dropout	classification of diabetic retinopathy
[68]	2021	Diabetes	MC-dropout	Segmentation and UQ of diabetic retinopathy lesions

sections. Medical image processing has gotten most of the attention regarding BDL in healthcare, whereas other tasks have received less devotion from researchers. Overall, image classification and segmentation have the largest number of published papers using BDL compared to other tasks, as demonstrated in section 6.1. Interestingly, the MC-Dropout was the most common method being used for medical image processing, while the VI method gained popularity among non-image tasks in healthcare, as illustrated in Figures 8 and 9. It is not surprising that MCMC has not gained much popularity. However, with less than a tenth of the work used, it deserves much more attention and investigation because it has the potential to overtake many techniques currently in use. From both the Bayesian and healthcare perspectives, this section will discuss the challenges and limitations of BDL methods in healthcare. Furthermore, future initiatives and some research gaps are discussed.

A. CHALLENGES FROM BDL PERSPECTIVE

Deep neural networks generally have more complex architectures than other machine learning techniques such as Decision Tree, SVM, and KNN. Despite having a complex architecture, which requires more computational power to train models, DL models became popular in the last decade because they achieved state-of-the-art results in numerous machine-learning tasks. This all became possible thanks to the availability of high computational power of computers. Comparing BDL to classical DL is somehow similar to comparing classical DL to other machine learning techniques. In several applications, BDL has proved to be superior to classical DL in several situations, particularly in situations where data has a high noise ratio or small datasets are used [54], [57], and [74]. However, BDL models are more complex

than classical DL and thus need more computational power to achieve high performance when training large models. The sampling process of BDL when sampling from the posterior distribution necessitates substantial computer capacity. For instance, when a model needs to find the precise posterior distribution of a specific parameter in a model's "weight", MCMC techniques should be used to sample from the posterior distribution, which might need thousands to millions of samples to fine-tune a single parameter in a given model [70], [72], [91]. For large models, this will be far from possible; hence, approximation methods have become popular in BDL applications. Approximation techniques have the disadvantage of not being as exact as MCMC sampling techniques. However, approximation techniques made it possible to apply BDL to some very large and complex models. Despite the scalability of approximation methods, models may require additional parameters and more training to reach equilibrium. When compared to classical DL, using VI for sampling with a Gaussian prior requires twice the number of parameters and more training for the model to converge [8].

Another challenge in BDL is the prior probability distribution, which is also known as the prior and is the expert's belief about the data and model parameters (weights) before seeing the data. This belief is the main difference between Bayesian and frequentist estimations in statistics. Prior knowledge in Bayesian models does take the shape of a probability distribution and it is based on an expert's knowledge, which can be represented in the model using different probability distributions such as Gaussian, Binomial, Gamma, Beta, Poisson, and others. Although the BDL has not been widespread among deep learning researchers, from the published works, researchers have rarely given importance to the use of different priors, which is the main purpose of using Bayesian

inference. Among the reviewed papers, only references [94] and [109] used and described different types of prior distributions for the unknown weights of the DL models. The reason that various prior distributions, other than Bernoulli for dropout and normal or Gaussian for VI, are seldom used is that most published works use approximation methods where normal/Gaussian works well for general purpose approximation. Another reason is that having some meaningful priors for all parameters (weights) in large complicated models with thousands or millions of parameters is extremely difficult, if not impossible. Despite the above-mentioned reasons, some probability distributions may fit better for some applications than others. Medical image models, for example, may perform better with normal or Gaussian priors, as illustrated in section 6.1, whereas the prediction for spreading COVID-19 will perform better with the Poisson distribution [109]. Further investigation is needed to come up with the best prior distribution choice for each type of application and data. For small-size datasets, it is better to investigate different prior distributions for each feature, in addition to MCMC sampling techniques to sample from the posterior distribution to benefit from the essence of the Bayesian model [94], [109].

The availability of proper tools to design and implement BDL models is essential to advancing in the field. In addition, open-source and other shared codes will have a vital role in advancing and investigating different aspects of BDL. A few frameworks have been developed for the implementation of BDL in the last decade, such as TensorFlow Probability, Pyro, PyMC3, Edward, and Stan. Compared to a few years ago, some of these libraries and frameworks have significantly improved, and developing models has become an easy task. However, compared to what Bayesian statistics covers, these tools are far from being enough or perfect for the task. This is because these tools only cover the basics of what Bayesian statistics offer, and there is a long way ahead for these tools to cover all aspects of Bayesian statistics, even though realistically not every bit of detail needs to be addressed by these tools. In addition to available tools, there is a lack of available code implementations of published works in BDL. For instance, based on our search of reviewed papers covered in the previous sections, only a limited number, which is roughly 20%, have made their code available, such as in references [45], [54], [56], [63], and [76], and some of the papers have not even described hyperparameters of their proposed models. We believe that providing source code and data and making them publicly available would considerably improve the pace and progress of BDL models.

B. CHALLENGES FROM A HEALTHCARE PERSPECTIVE

When a substantial amount of data is available, classical deep learning models tend to perform well and provide. However, Bayesian deep learning models are the exception, because they can perform well even when there are fewer data available. This is the case in healthcare applications, where sometimes only a limited number of data points are available, and in some cases, most available medical datasets [54], [57] are

small to medium in size compared to other fields of research. Therefore, having a large amount of data in healthcare is difficult for different reasons, including privacy issues and a lack of access to appropriate medical centers and facilities, especially in developing countries. The data availability issue can be partially resolved through the collaboration of different institutions involved in data collection and processing.

Another challenge with healthcare data is the lack of standards in its collection. Most of the time in healthcare, data collection might be done months or years apart by different medical specialists and different healthcare centers, resulting in data inconsistency and discrepancy. Data collected may include a substantial amount of noise, as well as redundant and missing data. Cleaning and preprocessing these types of inconsistencies in data is not an easy task, which might affect the quality of the data collected. Therefore, it is essential to have a standard procedure for data collection, and most of the time, there are some standard operating procedures in place. Yet, because of the difference between the available tools for data collection under different circumstances, data is not always collected uniformly. Another situation is when data is collected from already existing sources and different procedures are used during the data collection process.

Another main challenge in healthcare data is the distribution shift [45], [76]. The term “distribution shift” refers to the vast variation in data collected from different sources or at different times. For example, X-ray images could be collected from a few different hospitals, each of which uses a different type of machine with different settings. This issue is not easy to resolve since it may be very costly and time-consuming. Despite its good performance with noisy data, BDL cannot work its magic on distributionally shifted data where training and testing data are collected from different sources with different settings. This issue is what makes deploying DL models in healthcare so difficult since models underachieve in performance with new data. This issue can be alleviated slightly by employing techniques such as active learning, in which human intervention can help train models for new data, and incremental learning, in which models learn about new data without forgetting or retraining on old data.

Imbalance in training data is also a common issue in deep learning in general, and particularly in healthcare data [46], [55], [95], [99], and [110]. This problem particularly affects results in classification and regression models where data are not uniformly distributed. Models tend to perform very well on classes with a high occurrence ratio (majority class), while classes with a low occurrence ratio (minority class) may rarely be classified or predicted. Numerous methods have been developed to overcome data imbalance issues that improve the performance of models to a certain level. In general, BDL models perform better than classical DL when data is skewed and imbalanced to a certain level. However, for extreme cases of imbalanced data, which is the case in many healthcare datasets, these methods do not produce good results.

C. RESEARCH GAPS

In this section, we will go through some of the research gaps that have been raised in the implementation of BDL's application in healthcare. This list of research directions either has shortages in published work or has not yet been addressed by any means using BDL.

1) UNCERTAINTY QUANTIFICATION, INTERPRETABILITY, AND CAUSALITY

When researchers in the field discuss uncertainty in DL models, the first thing that comes to mind is the use of BDL, despite the fact that there are other alternative techniques for quantifying uncertainty, such as ensemble models and fuzzy logic. BDL has proven to be one of the most effective techniques for uncertainty quantification. When we go a level higher for clarification of model prediction, we will reach interpretability. Classical deep learning works as a black box since it cannot be determined how and why a given predicted output has been chosen based on a given input. Luckily, BDL can give insight on how and why a model predicted a certain output given some inputs thanks to the use of BDL structures. However, only a limited number of researchers have taken advantage of this functionality by using BDL, and from the reviewed papers in this study, we noticed that only Hua *et al.* [94] used BDL for such a task. Therefore, we believe that more research is needed in this area to further investigate the ability of BDL for interpretation tasks. Another fascinating topic in healthcare is causality, which has received little attention from researchers. Causality (also known as causal inference or casual analysis) is the study of finding the cause and effect of a certain outcome, which can be considered an optimal tool for analyzing healthcare data. Bayesian inference can be used as a technique for causal analysis. However, to the best of the authors' knowledge, no single published work in healthcare has used Bayesian inference for causality analysis for DL models. It has only been used by a few researchers in other areas of research. As a result, further study and investigation are required in this field to offer knowledge of how valuable this field can be in healthcare applications.

2) GENERAL GAPS OF BDL IN HEALTHCARE

Earlier sections of this study have reviewed BDL in healthcare, such as medical imaging and signal processing. Although some of these papers have been reviewed, the majority of published work has only covered the surface of topics and opportunities in the field. In medical imaging, for example, some papers have been published on the classification task. Yet, there is a lot to be investigated when it comes to classification because most of the published work has only covered multiclass classification tasks. Nonetheless, there are other classification types, such as binary classification, multi-label classification, and hierarchical classification, where BDL has not been used to model such problems and quantify model uncertainty. Another task for medical image analysis is segmentation, where only a few types of image segmentation

have been covered by published work using BDL. To determine the practicality of these models and how the UQ affects performance for different types of segmentation, such as semantic segmentation, object/instance segmentation, panoptic segmentation, and edge segmentation, all types of image segmentation should be investigated with BDL. Other medical image tasks, such as image registration, image super-resolution, and image reconstruction, have likewise received little attention. To the best of the authors' knowledge, there is no single published work using BDL in clinical natural language processing, which might create opportunities for researchers to work in the field. EHR, audio, and video processing have barely taken advantage of BDL, making them an open subject for BDL to further investigate in more detail, especially for uncertainty quantification tasks. In addition, due to the lack of published work in those areas, such as medical image captioning, medical image to a sequence that can describe details of images, and text generation for clinical reports, there are some areas of research that could have a huge benefit in healthcare, as described in previous sections.

3) GAPS IN USING BDL FOR DIFFERENT MACHINE LEARNING TASKS IN HEALTHCARE

Most of the published works in healthcare using BDL have been reviewed in the previous sections. Among those, only a few machine learning tasks have been covered, and some of them have only a few related published works. It is essential to know how much BDL can influence each of the remaining machine learning tasks and deep learning architectures. Despite its significant importance in healthcare, semi-supervised learning is one of the useful machine learning tasks that has not received the attention it deserves. Semi-supervised learning comes between supervised and unsupervised learning, where only a small amount of data has labels for model training. Among the reviewed papers, only Sedai *et al.* [61] used BDL for semi-supervised learning for the segmentation task of the retinal layer. Another task that has a shortage of using BDL for healthcare tasks is active learning. Active learning requires human intervention to label data as needed, since the dataset may not have enough labeled data. This task is also important for healthcare applications since medical data labels do not always exist and the process of labeling may be costly and time-consuming. Only a few papers, including Rączkowski *et al.* [43], Gal *et al.* [83], Mahapatra *et al.* [84], Saidu *et al.* [60], and Bodenstedt *et al.* [32], have used BDL for active learning tasks. Furthermore, except for Hassan *et al.*'s [82] work, continual learning and incremental learning are two different machine learning tasks that can benefit healthcare but have not been sufficiently investigated with BDL. In continual learning, data is fed continuously to the model for training, while in incremental learning; a trained model is retrained on new data without having to use old data for training. Another intriguing area that has received little attention is Graph neural networks, which can also be used to aid with graph data in healthcare and benefit from BDL for uncertainty estimation.

No published work has used a Bayesian graph neural network for healthcare data, which opens up new opportunities in this field.

IX. CONCLUSION

Deep learning has scaled up the use of machine learning to an unprecedented level and has been deployed in most fields such as computer vision, natural language processing, and signal processing. In healthcare applications, deep learning models have been intensively used for medical imaging tasks, signal processing, and electronic health record analysis. Despite their success, classical deep learning models are suffering from some issues, such as overfitting and establishing the confidence level of the models' output. To overcome these issues, BDL has developed in recent years. For healthcare data in general, BDL can be a better choice since it can determine the uncertainty associated with data and is less susceptible to noisy data, making it a good choice as a model regularization technique. This paper provides a comprehensive review of common BDL models used in the healthcare field. Additionally, the most common Bayesian inference methods in deep learning, such as MCMC, VI, and MC-dropout, are described and reviewed. This paper reviewed most of the published works that employed BDL for medical imaging tasks such as image classification, segmentation, registration, reconstruction, and enhancement. Other areas in healthcare have also been reviewed, such as medical signal processing, electronic health records, and video processing, along with the implementation of BDL in a few widespread diseases. Moreover, we discussed some challenges facing researchers in both healthcare applications and BDL implementation, ending by presenting some future research gaps and directions in this field.

REFERENCES

- [1] J. Huertas-Tato, A. Martín, J. Fierrez, and D. Camacho, "Fusing CNNs and statistical indicators to improve image classification," *Inf. Fusion*, vol. 79, pp. 174–187, Mar. 2022, doi: [10.1016/j.inffus.2021.09.012](https://doi.org/10.1016/j.inffus.2021.09.012).
- [2] P. Teare, M. Fishman, O. Benzaquen, E. Toledano, and E. Elnekave, "Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement," *J. Digit. Imag.*, vol. 30, no. 4, pp. 499–505, Aug. 2017, doi: [10.1007/s10278-017-9993-2](https://doi.org/10.1007/s10278-017-9993-2).
- [3] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The rise of deep learning in drug discovery," *Drug Discovery Today*, vol. 23, no. 6, pp. 1241–1250, Jun. 2018, doi: [10.1016/j.drudis.2018.01.039](https://doi.org/10.1016/j.drudis.2018.01.039).
- [4] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, and O. Geessink, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *J. Amer. Med. Assoc.*, vol. 318, no. 22, p. 2199, Dec. 2017, doi: [10.1001/jama.2017.14585](https://doi.org/10.1001/jama.2017.14585).
- [5] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020, doi: [10.3390/app10093233](https://doi.org/10.3390/app10093233).
- [6] S. Tuli, N. Basumatary, S. S. Gill, M. Kahani, R. C. Arya, G. S. Wander, and R. Buyya, "HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments," *Future Gener. Comput. Syst.*, vol. 104, pp. 187–200, Mar. 2020, doi: [10.1016/j.future.2019.10.043](https://doi.org/10.1016/j.future.2019.10.043).
- [7] X. Zhou, W. Liang, K. I.-K. Wang, H. Wang, L. T. Yang, and Q. Jin, "Deep-learning-enhanced human activity recognition for internet of healthcare things," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6429–6438, Jul. 2020, doi: [10.1109/JIOT.2020.2985082](https://doi.org/10.1109/JIOT.2020.2985082).
- [8] O. Dürr, B. Sick, and E. Murina, *Probabilistic Deep Learning: With Python, Keras and TensorFlow Probability*. Shelter Island, NY, USA: Manning Publications Company, 2020.
- [9] S. Shamsirband, M. Fathi, A. Dehzaangi, A. T. Chronopoulos, and H. Alinejad-Rokny, "A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues," *J. Biomed. Informat.*, vol. 113, Jan. 2021, Art. no. 103627, doi: [10.1016/j.jbi.2020.103627](https://doi.org/10.1016/j.jbi.2020.103627).
- [10] D.-E.-M. Nisar, R. Amin, N.-U.-H. Shah, M. A. A. Ghamdi, S. H. Almotiri, and M. Alruily, "Healthcare techniques through deep learning: Issues, challenges and opportunities," *IEEE Access*, vol. 9, pp. 98523–98541, 2021, doi: [10.1109/ACCESS.2021.3095312](https://doi.org/10.1109/ACCESS.2021.3095312).
- [11] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 156–180, 2021, doi: [10.1109/RBME.2020.3013489](https://doi.org/10.1109/RBME.2020.3013489).
- [12] S. K. Pandey and R. R. Janghel, "Recent deep learning techniques, challenges and its applications for medical healthcare system: A review," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1907–1935, Oct. 2019, doi: [10.1007/s11063-018-09976-2](https://doi.org/10.1007/s11063-018-09976-2).
- [13] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H.-J. Bae, and N. Kim, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, pp. 657–668, Dec. 2019, doi: [10.14245/ns.1938396.198](https://doi.org/10.14245/ns.1938396.198).
- [14] N. G. Polson and V. Sokolov, "Deep learning: A Bayesian perspective," *Bayesian Anal.*, vol. 12, no. 4, pp. 1275–1304, Dec. 2017, doi: [10.1214/17-BA1082](https://doi.org/10.1214/17-BA1082).
- [15] H. Wang and D.-Y. Yeung, "Towards Bayesian deep learning: A framework and some existing methods," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3395–3408, Dec. 2016, doi: [10.1109/TKDE.2016.2606428](https://doi.org/10.1109/TKDE.2016.2606428).
- [16] 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](https://doi.org/10.1145/3409383).
- [17] J. Xuan, J. Lu, and G. Zhang, "A survey on Bayesian nonparametric learning," *ACM Comput. Surv.*, vol. 52, no. 1, pp. 1–36, Jan. 2020, doi: [10.1145/3291044](https://doi.org/10.1145/3291044).
- [18] T. Charnock, L. Perreault-Levasseur, and F. Lanusse, "Bayesian neural networks," in *Artificial Intelligence for High Energy Physics*. Singapore: World Scientific, 2022, pp. 663–713, doi: [10.1142/9789811234033_0018](https://doi.org/10.1142/9789811234033_0018).
- [19] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021, doi: [10.1016/j.inffus.2021.05.008](https://doi.org/10.1016/j.inffus.2021.05.008).
- [20] R. Alizadehsani, M. Roshanzamir, S. Hussain, A. Khosravi, A. Koohestani, M. H. Zangoeei, M. Abdar, A. Beykikhoshk, A. Shoeibi, A. Zare, M. Panahiazar, S. Nahavandi, D. Srinivasan, A. F. Atiya, and U. R. Acharya, "Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020)," *Ann. Operations Res.*, Mar. 2021, doi: [10.1007/s10479-021-04006-2](https://doi.org/10.1007/s10479-021-04006-2).
- [21] S. Depeweg, J. M. Hernandez-Lobato, F. Doshi-Velez, and S. Udluft, "Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, 2018, pp. 1–10.
- [22] A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?" *Struct. Saf.*, vol. 31, no. 2, pp. 105–112, Mar. 2009, doi: [10.1016/j.strusafe.2008.06.020](https://doi.org/10.1016/j.strusafe.2008.06.020).
- [23] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1–10.
- [24] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds Mach.*, vol. 30, no. 4, pp. 681–694, Dec. 2020, doi: [10.1007/s11023-020-09548-1](https://doi.org/10.1007/s11023-020-09548-1).
- [25] A. Bate, M. Lindquist, I. R. Edwards, S. Olsson, R. Orre, A. Lansner, and R. M. De Freitas, "A Bayesian neural network method for adverse drug reaction signal generation," *Eur. J. Clin. Pharmacol.*, vol. 54, no. 4, pp. 315–321, Jul. 1998, doi: [10.1007/s002280050466](https://doi.org/10.1007/s002280050466).

- [26] B. Baesens, S. Viaene, D. van den Poel, J. Vanthienen, and G. Dedene, "Bayesian neural network learning for repeat purchase modelling in direct marketing," *Eur. J. Oper. Res.*, vol. 138, no. 1, pp. 191–211, 2002, doi: [10.1016/S0377-2217\(01\)00129-1](https://doi.org/10.1016/S0377-2217(01)00129-1).
- [27] A. Géron, *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA, USA: O'Reilly, 2017.
- [28] F. Chollet, *Deep Learning With Python*, vol. 53, no. 9. Shelter Island, NY, USA: Manning, 2019.
- [29] 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](https://doi.org/10.1016/j.neuroimage.2018.03.065).
- [30] Y. Tang and X. Wu, "Text-independent writer identification via CNN features and joint Bayesian," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 566–571, doi: [10.1109/ICFHR.2016.0109](https://doi.org/10.1109/ICFHR.2016.0109).
- [31] P. Waldmann, "Approximate Bayesian neural networks in genomic prediction," *Genet. Selection Evol.*, vol. 50, no. 1, pp. 1–9, Dec. 2018, doi: [10.1186/s12711-018-0439-1](https://doi.org/10.1186/s12711-018-0439-1).
- [32] S. Bodenstedt, D. Rivoir, A. Jenke, M. Wagner, M. Breucha, B. Müller-Stich, S. T. Mees, J. Weitz, and S. Speidel, "Active learning using deep Bayesian networks for surgical workflow analysis," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 6, pp. 1079–1087, Jun. 2019, doi: [10.1007/s11548-019-01963-9](https://doi.org/10.1007/s11548-019-01963-9).
- [33] J.-T. Chien and Y.-C. Ku, "Bayesian recurrent neural network for language modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 361–374, Feb. 2016, doi: [10.1109/TNNLS.2015.2499302](https://doi.org/10.1109/TNNLS.2015.2499302).
- [34] P. McDermott and C. Wikle, "Bayesian recurrent neural network models for forecasting and quantifying uncertainty in spatial-temporal data," *Entropy*, vol. 21, no. 2, p. 184, Feb. 2019, doi: [10.3390/e21020184](https://doi.org/10.3390/e21020184).
- [35] J. Albert and J. Hu, *Probability and Bayesian Modeling*. Boca Raton, FL, USA: CRC Press, 2019, doi: [10.1201/9781351030144](https://doi.org/10.1201/9781351030144).
- [36] F. Liang, C. Liu, and R. J. Carroll, *Advanced Markov Chain Monte Carlo Methods Learning From Past Samples*. Hoboken, NJ, USA: Wiley, 2010, doi: [10.1002/9780470669723](https://doi.org/10.1002/9780470669723).
- [37] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, vol. 1, 2nd ed. New York, NY, USA: Texts Stat. Sci., 2006.
- [38] R. Mubvha, W. T. Mongwe, and T. Marwala, "Separable shadow Hamiltonian hybrid Monte Carlo for Bayesian neural network inference in wind speed forecasting," *Energy AI*, vol. 6, Dec. 2021, Art. no. 100108, doi: [10.1016/j.egyai.2021.100108](https://doi.org/10.1016/j.egyai.2021.100108).
- [39] R. Chandra, A. Bhagat, M. Maharana, and P. N. Krivitsky, "Bayesian graph convolutional neural networks via tempered MCMC," *IEEE Access*, vol. 9, pp. 130353–130365, 2021, doi: [10.1109/ACCESS.2021.3111898](https://doi.org/10.1109/ACCESS.2021.3111898).
- [40] A. Wu, S. Nowozin, E. Meeds, R. E. Turner, J. M. Hernández-Lobato, and A. L. Gaunt, "Deterministic variational inference for robust Bayesian neural networks," 2019, *arXiv:1810.03958*.
- [41] M. Haußmann, F. A. Hamprecht, and M. Kandemir, "Sampling-free variational inference of Bayesian neural networks by variance back-propagation," in *Uncertainty in Artificial Intelligence*. PMLR, 2019, pp. 563–573.
- [42] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jan. 2014.
- [43] Ł. Rączkowski, M. Możejko, J. Zambonelli, and E. Szczurek, "ARA: Accurate, reliable and active histopathological image classification framework with Bayesian deep learning," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019, doi: [10.1038/s41598-019-50587-1](https://doi.org/10.1038/s41598-019-50587-1).
- [44] B. Song, S. Sunny, S. Li, K. Gurushanth, P. Mendonca, N. Mukhia, S. Patrick, S. Gurudath, S. Raghavan, I. Tsusennaro, and S. T. Leivon, "Bayesian deep learning for reliable oral cancer image classification," *Biomed. Opt. Exp.*, vol. 12, no. 10, p. 6422, Oct. 2021, doi: [10.1364/boe.432365](https://doi.org/10.1364/boe.432365).
- [45] A. Filos, S. Farquhar, A. N. Gomez, T. G. J. Rudner, Z. Kenton, L. Smith, M. Alizadeh, A. de Kroon, and Y. Gal, "A systematic comparison of Bayesian deep learning robustness in diabetic retinopathy tasks," 2019, *arXiv:1912.10481*.
- [46] S. Yadav, "Bayesian deep learning based convolutional neural network for classification of Parkinson's disease using functional magnetic resonance images," *SSRN Electron. J.*, pp. 1–22, 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3833760, doi: [10.2139/ssrn.3833760](https://doi.org/10.2139/ssrn.3833760).
- [47] Y. Liu, C. Zhao, and J. Rubin, "Uncertainty quantification in chest X-ray image classification using Bayesian deep neural networks," in *Proc. CEUR Workshop*, vol. 2675, 2020, pp. 19–26.
- [48] W. J. Maddox, P. Izmailov, T. Garipov, D. P. Vetrov, and A. G. Wilson, "A simple baseline for Bayesian uncertainty in deep learning," in *Proc. Adv. in Neural Inf. Process. Syst. (NeurIPS)*, vol. 32, 2019.
- [49] M. H.-M. Khan, "Automated breast cancer diagnosis using artificial neural network (ANN)," in *Proc. 3rd Iranian Conf. Intell. Syst. Signal Process. (ICSPIS)*, Dec. 2017, pp. 54–58, doi: [10.1109/ICSPIS.2017.8311589](https://doi.org/10.1109/ICSPIS.2017.8311589).
- [50] 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. Statist. Data Anal.*, vol. 142, Feb. 2020, Art. no. 106816, doi: [10.1016/j.csda.2019.106816](https://doi.org/10.1016/j.csda.2019.106816).
- [51] P. Thiagarajan, P. Khairnar, and S. Ghosh, "Explanation and use of uncertainty obtained by Bayesian neural network classifiers for breast histopathology images," *IEEE Trans. Med. Imag.*, early access, Oct. 26, 2021, doi: [10.1109/TMI.2021.3123300](https://doi.org/10.1109/TMI.2021.3123300).
- [52] P. van Molle, T. Verbelen, C. De Boom, B. Vankeirsbilck, J. De Vylder, B. Diricx, T. Kimpe, P. Simoons, and B. Dhoedt, "Quantifying uncertainty of deep neural networks in skin lesion classification," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures* (Lecture Notes in Computer Science), vol. 11840. Cham, Switzerland: Springer, 2019, doi: [10.1007/978-3-030-32689-0_6](https://doi.org/10.1007/978-3-030-32689-0_6).
- [53] I. Oloyede, "Bayesian deep learning algorithms for classifying COVID-19 images," *Int. J. Comput. Inf. Eng.*, vol. 15, no. 2, pp. 145–149, Jan. 2021.
- [54] M. Gour and S. Jain, "Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification," *Comput. Biol. Med.*, vol. 140, Jan. 2022, Art. no. 105047, doi: [10.1016/j.combiomed.2021.105047](https://doi.org/10.1016/j.combiomed.2021.105047).
- [55] J. I. Orlando, P. Seebeck, H. Bogunovic, S. Klimescha, C. Grechenig, S. Waldstein, B. S. Gerendas, and U. Schmidt-Erfurth, "U2-Net: A Bayesian U-Net model with epistemic uncertainty feedback for photoreceptor layer segmentation in pathological OCT scans," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1441–1445, doi: [10.1109/ISBI.2019.8759581](https://doi.org/10.1109/ISBI.2019.8759581).
- [56] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control," *NeuroImage*, vol. 195, pp. 11–22, Jul. 2019, doi: [10.1016/j.neuroimage.2019.03.042](https://doi.org/10.1016/j.neuroimage.2019.03.042).
- [57] P. McClure, N. Rho, J. A. Lee, J. R. Kaczmarzyk, C. Y. Zheng, S. S. Ghosh, D. M. Nielson, A. G. Thomas, P. Bandettini, and F. Pereira, "Knowing what you know in brain segmentation using Bayesian deep neural networks," *Frontiers Neuroinform.*, vol. 13, p. 67, Oct. 2019, doi: [10.3389/fninf.2019.00067](https://doi.org/10.3389/fninf.2019.00067).
- [58] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical CT using Bayesian U-Net for personalized musculoskeletal modeling," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1030–1040, Apr. 2020, doi: [10.1109/TMI.2019.2940555](https://doi.org/10.1109/TMI.2019.2940555).
- [59] J. Ma, F. Lin, S. Wesarg, and M. Erdt, "A novel Bayesian model incorporating deep neural network and statistical shape model for pancreas segmentation," in *Proc. Int. Conf. Med. Image Comput.-Assist. Intervent.*, 2018, pp. 480–487, doi: [10.1007/978-3-030-00937-3_55](https://doi.org/10.1007/978-3-030-00937-3_55).
- [60] I. C. Saidu and L. Csató, "Active learning with Bayesian UNet for efficient semantic image segmentation," *J. Imag.*, vol. 7, no. 2, p. 37, Feb. 2021, doi: [10.3390/jimaging7020037](https://doi.org/10.3390/jimaging7020037).
- [61] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, and R. Garnavi, "Uncertainty guided semi-supervised segmentation of retinal layers in OCT images," *J. Imag.*, vol. 7, no. 2, 2019, doi: [10.1007/978-3-030-32239-7_32](https://doi.org/10.1007/978-3-030-32239-7_32).
- [62] J. Sander, B. D. de Vos, J. M. Wolterink, and I. Išgum, "Towards increased trustworthiness of deep learning segmentation methods on cardiac MRI," in *Proc. Med. Imag., Image Process.*, vol. 10949. Bellingham, WA, USA: SPIE, 2019, pp. 324–330, doi: [10.1117/12.2511699](https://doi.org/10.1117/12.2511699).
- [63] S. Sedai, B. Antony, D. Mahapatra, and R. Garnavi, "Joint segmentation and uncertainty visualization of retinal layers in optical coherence tomography images using Bayesian deep learning," in *Computational Pathology and Ophthalmic Medical Image Analysis* (Lecture Notes in Computer Science), vol. 11039. Cham, Switzerland: Springer, 2018, pp. 219–227, doi: [10.1007/978-3-030-00949-6_26](https://doi.org/10.1007/978-3-030-00949-6_26).

- [64] R. Jena and S. P. Awate, "A Bayesian neural net to segment images with uncertainty estimates and good calibration," in *Information Processing in Medical Imaging* (Lecture Notes in Computer Science) vol. 11492. Cham, Switzerland: Springer, 2019, pp. 3–15, doi: [10.1007/978-3-030-20351-1_1](https://doi.org/10.1007/978-3-030-20351-1_1).
- [65] M. Antico, F. Sasazawa, Y. Takeda, A. T. Jaiprakash, M.-L. Wille, A. K. Pandey, R. Crawford, G. Carneiro, and D. Fontanarosa, "Bayesian CNN for segmentation uncertainty inference on 4D ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy," *IEEE Access*, vol. 8, pp. 223961–223975, 2020, doi: [10.1109/ACCESS.2020.3044355](https://doi.org/10.1109/ACCESS.2020.3044355).
- [66] Y. Liu, G. Zhao, B. M. Nacewicz, N. Adluru, G. R. Kirk, P. A. Ferrazzano, M. Styner, and A. L. Alexander, "Accurate automatic segmentation of amygdala subnuclei and modeling of uncertainty via Bayesian fully convolutional neural network," Feb. 2019, *arXiv:1902.07289*.
- [67] Y. Liu, G. Yang, M. Hosseiny, A. Azadikhah, S. A. Mirak, Q. Miao, S. S. Raman, and K. Sung, "Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation," *IEEE Access*, vol. 8, pp. 151817–151828, 2020, doi: [10.1109/ACCESS.2020.3017168](https://doi.org/10.1109/ACCESS.2020.3017168).
- [68] A. Garifullin, L. Lensu, and H. Uusitalo, "Deep Bayesian baseline for segmenting diabetic retinopathy lesions: Advances and challenges," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104725, doi: [10.1016/j.cmbiomed.2021.104725](https://doi.org/10.1016/j.cmbiomed.2021.104725).
- [69] A. Largent, J. De Asis-Cruz, K. Kapse, S. D. Barnett, J. Murnick, S. Basu, N. Andersen, S. Norman, N. Andescavage, and C. Limperopoulos, "Automatic brain segmentation in preterm infants with post-hemorrhagic hydrocephalus using 3D Bayesian U-Net," *Human Brain Mapping*, vol. 43, no. 6, pp. 1895–1916, Apr. 2022, doi: [10.1002/hbm.25762](https://doi.org/10.1002/hbm.25762).
- [70] L. L. Folgoc, H. Delingette, A. Criminisi, and N. Ayache, "Quantifying registration uncertainty with sparse Bayesian modelling," *IEEE Trans. Med. Imag.*, vol. 36, no. 2, pp. 607–617, Feb. 2017, doi: [10.1109/TMI.2016.2623608](https://doi.org/10.1109/TMI.2016.2623608).
- [71] V. S. Deshpande and J. S. Bhatt, "Bayesian deep learning for deformable medical image registration," in *Pattern Recognition and Machine Intelligence* (Lecture Notes in Computer Science), vol. 11942. Cham, Switzerland: Springer, 2019, doi: [10.1007/978-3-030-34872-4_5](https://doi.org/10.1007/978-3-030-34872-4_5).
- [72] S. Khawaled and M. Freiman, "Unsupervised deep-learning based deformable image registration: A Bayesian framework," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.* Cham, Switzerland: Springer, Aug. 2020, pp. 41–49.
- [73] K. Cui, P. Fu, Y. Li, and Y. Lin, "Bayesian fully convolutional networks for brain image registration," *J. Healthcare Eng.*, vol. 2021, pp. 1–12, Jul. 2021, doi: [10.1155/2021/5528160](https://doi.org/10.1155/2021/5528160).
- [74] J. Schlemper, D. C. Castro, W. Bai, C. Qin, O. Oktay, J. Duan, A. N. Price, J. Hajnal, and D. Rueckert, "Bayesian deep learning for accelerated mr image reconstruction," in *Machine Learning for Medical Image Reconstruction* (Lecture Notes in Computer Science), vol. 11074. Cham, Switzerland: Springer, 2018, pp. 64–71, doi: [10.1007/978-3-030-00129-2_8](https://doi.org/10.1007/978-3-030-00129-2_8).
- [75] C. Du, C. Du, L. Huang, and H. He, "Reconstructing perceived images from human brain activities with Bayesian deep multiview learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 8, pp. 2310–2323, Aug. 2019, doi: [10.1109/TNNLS.2018.2882456](https://doi.org/10.1109/TNNLS.2018.2882456).
- [76] R. V. Marinescu, D. Moyer, and P. Golland, "Bayesian image reconstruction using deep generative models," Dec. 2020, *arXiv:2012.04567*.
- [77] B. Yang, L. Ying, and J. Tang, "Artificial neural network enhanced Bayesian PET image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1297–1309, Jun. 2018, doi: [10.1109/TMI.2018.2803681](https://doi.org/10.1109/TMI.2018.2803681).
- [78] R. Barbano, Z. Kereta, A. Hauptmann, S. R. Arridge, and B. Jin, "Unsupervised knowledge-transfer for learned image reconstruction," Jul. 2021, *arXiv:2107.02572*.
- [79] R. Tanno, D. E. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Uncertainty modelling in deep learning for safer NeuroImage enhancement: Demonstration in diffusion MRI," *NeuroImage*, vol. 225, Jan. 2021, Art. no. 117366, doi: [10.1016/j.neuroimage.2020.117366](https://doi.org/10.1016/j.neuroimage.2020.117366).
- [80] R. Tanno, D. E. Worrall, A. Ghosh, E. Kaden, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander, "Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution," in *Medical Image Computing and Computer Assisted Intervention* (Lecture Notes in Computer Science), vol. 10433. 2017, doi: [10.1007/978-3-319-66182-7_70](https://doi.org/10.1007/978-3-319-66182-7_70).
- [81] Y. Li, F. Xu, F. Zhang, P. Xu, M. Zhang, M. Fan, L. Li, X. Gao, and R. Han, "DLBI: Deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy," *Bioinformatics*, vol. 34, no. 13, pp. i284–i294, Jul. 2018, doi: [10.1093/bioinformatics/bty241](https://doi.org/10.1093/bioinformatics/bty241).
- [82] T. Hassan, B. Hassan, M. U. Akram, S. Hashmi, A. H. Taguri, and N. Werghi, "Incremental cross-domain adaptation for robust retinopathy screening via Bayesian deep learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021, doi: [10.1109/TIM.2021.3122172](https://doi.org/10.1109/TIM.2021.3122172).
- [83] Y. Gal, R. Islam, and Z. Ghahramani, "Deep Bayesian active learning with image data," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 1183–1192, doi: [10.17863/CAM.11070](https://doi.org/10.17863/CAM.11070).
- [84] D. Mahapatra, B. Bozorgtabar, J. P. Thiran, and M. Reyes, "Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network," in *Medical Image Computing and Computer Assisted Intervention* (Lecture Notes in Computer Science), vol. 11071. Cham, Switzerland: Springer, 2018, pp. 580–588, doi: [10.1007/978-3-030-00934-2_65](https://doi.org/10.1007/978-3-030-00934-2_65).
- [85] X. Gou and X. He, "Deep learning-based detection and diagnosis of subarachnoid hemorrhage," *J. Healthcare Eng.*, vol. 2021, pp. 1–10, Nov. 2021, doi: [10.1155/2021/9639419](https://doi.org/10.1155/2021/9639419).
- [86] T. Nakao, S. Hanaoka, Y. Nomura, N. Hayashi, and O. Abe, "Anomaly detection in chest 18F-FDG PET/CT by Bayesian deep learning," *Japanese J. Radiol.*, vol. 324, pp. 1–10, Jan. 2022, doi: [10.1007/s11604-022-01249-2](https://doi.org/10.1007/s11604-022-01249-2).
- [87] R. Chai, Y. Tran, G. R. Naik, T. N. Nguyen, S. H. Ling, A. Craig, and H. T. Nguyen, "Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, doi: [10.1109/EMBC.2016.7591765](https://doi.org/10.1109/EMBC.2016.7591765).
- [88] R. Chai, M. R. Smith, T. N. Nguyen, S. Ho Ling, A. J. Coutts, and H. T. Nguyen, "Comparing features extractors in EEG-based cognitive fatigue detection of demanding computer tasks," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, doi: [10.1109/EMBC.2015.7320150](https://doi.org/10.1109/EMBC.2015.7320150).
- [89] C. Q. Ngo, R. Chai, T. V. Nguyen, T. W. Jones, and H. T. Nguyen, "Nocturnal hypoglycemia detection using EEG spectral moments under natural occurrence conditions," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 7177–7180, doi: [10.1109/EMBC.2019.8856695](https://doi.org/10.1109/EMBC.2019.8856695).
- [90] A. M. A. Handojoseno, G. R. Naik, M. Gilat, J. M. Shine, T. N. Nguyen, Q. T. Ly, S. J. G. Lewis, and H. T. Nguyen, "Prediction of freezing of gait in patients with Parkinson's disease using EEG signals," *Studies Health Technol. Inform.*, vol. 246, pp. 124–131, 2018, doi: [10.3233/978-1-61499-845-7-124](https://doi.org/10.3233/978-1-61499-845-7-124).
- [91] W. Fruehwirt, A. D. Cobb, M. Mairhofer, L. Weydemann, H. Garn, R. Schmidt, T. Benke, P. Dal-Bianco, G. Ransmayr, M. Waser, and D. Grossegger, "Bayesian deep neural networks for low-cost neuro-physiological markers of Alzheimer's disease severity," Dec. 2018, *arXiv:1812.04994*.
- [92] S. S. Lee, K. Lee, and G. Kang, "EEG artifact removal by Bayesian deep learning ICA," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 932–935, doi: [10.1109/EMBC44109.2020.9175785](https://doi.org/10.1109/EMBC44109.2020.9175785).
- [93] A. O. Aseeri, "Uncertainty-aware deep learning-based cardiac arrhythmias classification model of electrocardiogram signals," *Computers*, vol. 10, no. 6, p. 82, Jun. 2021, doi: [10.3390/COMPUTERS10060082](https://doi.org/10.3390/COMPUTERS10060082).
- [94] Q. Hua, Y. Yaqin, B. Wan, B. Chen, Y. Zhong, and J. Pan, "An interpretable model for ECG data based on Bayesian neural networks," *IEEE Access*, vol. 9, pp. 57001–57009, 2021, doi: [10.1109/ACCESS.2021.3071731](https://doi.org/10.1109/ACCESS.2021.3071731).
- [95] D. J. Jagannath, D. R. J. Dolly, and J. D. Peter, "A novel Bayesian deep learning methodology for enhanced foetal cardiac signal mining," *J. Experim. Theor. Artif. Intell.*, vol. 31, no. 2, pp. 215–224, Mar. 2019, doi: [10.1080/0952813X.2018.1544283](https://doi.org/10.1080/0952813X.2018.1544283).
- [96] S. S. S. Das, S. K. Shanto, M. Rahman, M. Islam, A. Rahman, M. M. Masud, and M. E. Ali, "BayesBeat: A Bayesian deep learning approach for atrial fibrillation detection from noisy photoplethysmography data," Nov. 2020, *arXiv:2011.00753*.
- [97] J. Belen, S. Mousavi, A. Shamsoshoara, and F. Afghah, "An uncertainty estimation framework for risk assessment in deep learning-based AFib classification," in *Proc. 54th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2020, pp. 960–964, doi: [10.1109/IEEECONF51394.2020.9443466](https://doi.org/10.1109/IEEECONF51394.2020.9443466).

- [98] S. Bacchi, S. Gluck, Y. Tan, I. Chim, J. Cheng, T. Gilbert, D. K. Menon, J. Jannes, T. Kleinig, and S. Koblar, "Prediction of general medical admission length of stay with natural language processing and deep learning: A pilot study," *Internal Emergency Med.*, vol. 15, no. 6, pp. 989–995, Sep. 2020, doi: [10.1007/s11739-019-02265-3](https://doi.org/10.1007/s11739-019-02265-3).
- [99] L. Chen, Y. Gu, X. Ji, Z. Sun, H. Li, Y. Gao, and Y. Huang, "Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 1, pp. 56–64, Jan. 2020, doi: [10.1093/jamia/ocz141](https://doi.org/10.1093/jamia/ocz141).
- [100] M. C. Chen, R. L. Ball, L. Yang, N. Moradzadeh, B. E. Chapman, D. B. Larson, C. P. Langlotz, T. J. Amrhein, and M. P. Lungren, "Deep learning to classify radiology free-text reports," *Radiology*, vol. 286, no. 3, pp. 845–852, Mar. 2018, doi: [10.1148/radiol.2017171115](https://doi.org/10.1148/radiol.2017171115).
- [101] Y. Peng, K. Yan, V. Sandfort, R. M. Summers, and Z. Lu, "A self-attention based deep learning method for lesion attribute detection from CT reports," in *Proc. IEEE Int. Conf. Healthcare Inform. (ICHI)*, Jun. 2019, pp. 1–5, doi: [10.1109/ICHI.2019.8904668](https://doi.org/10.1109/ICHI.2019.8904668).
- [102] M. W. Dusenberry, D. Tran, E. Choi, J. Kemp, J. Nixon, G. Jerfel, K. Heller, and A. M. Dai, "Analyzing the role of model uncertainty for electronic health records," in *Proc. ACM Conf. Health, Inference, Learn.*, Apr. 2020, pp. 204–213, doi: [10.1145/3368555.3384457](https://doi.org/10.1145/3368555.3384457).
- [103] R. Qiu, Y. Jia, M. Hadzikadic, M. Dulin, X. Niu, and X. Wang, "Modeling the uncertainty in electronic health records: A Bayesian deep learning approach," Jul. 2019, *arXiv:1907.06162*.
- [104] J. Deasy, A. Ercole, and P. Liò, "Adaptive prediction timing for electronic health records," Mar. 2020, *arXiv:2003.02554*.
- [105] Y. Li, S. Rao, A. Hassaine, R. Ramakrishnan, D. Canoy, G. Salimi-Khorshidi, M. Mamouei, T. Lukasiewicz, and K. Rahimi, "Deep Bayesian Gaussian processes for uncertainty estimation in electronic health records," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, Dec. 2021, doi: [10.1038/s41598-021-00144-6](https://doi.org/10.1038/s41598-021-00144-6).
- [106] J. G. Diaz, O. L. Maier, and O. Csiszár, "Bayesian logical neural networks for human-centered applications in medicine," *medRxiv*, 2021, doi: [10.1101/2021.11.15.21266351](https://doi.org/10.1101/2021.11.15.21266351).
- [107] S. Liu, S. Hu, Y. Wang, J. Yu, R. Su, X. Liu, and H. Meng, "Exploiting visual features using Bayesian gated neural networks for disordered speech recognition," in *Proc. Interspeech*, Sep. 2019, pp. 4120–4124, doi: [10.21437/Interspeech.2019-1536](https://doi.org/10.21437/Interspeech.2019-1536).
- [108] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," Mar. 2020, *arXiv:2003.10769*.
- [109] S. Cabras, "A Bayesian-deep learning model for estimating COVID-19 evolution in Spain," *Mathematics*, vol. 9, no. 22, p. 2921, Nov. 2021, doi: [10.3390/math9222921](https://doi.org/10.3390/math9222921).
- [110] F. Ucar and D. Korkmaz, "COVIDdiagnosis-Net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images," *Med. Hypotheses*, vol. 140, Jul. 2020, Art. no. 109761, doi: [10.1016/j.mehy.2020.109761](https://doi.org/10.1016/j.mehy.2020.109761).
- [111] M. Loey, S. El-Sappagh, and S. Mirjalili, "Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data," *Comput. Biol. Med.*, vol. 142, Mar. 2022, Art. no. 105213, doi: [10.1016/j.compbiomed.2022.105213](https://doi.org/10.1016/j.compbiomed.2022.105213).
- [112] P. Niraula, J. Mateu, and S. Chaudhuri, "A Bayesian machine learning approach for spatio-temporal prediction of COVID-19 cases," *Stochastic Environ. Res. Risk Assessment*, vol. 534, pp. 1–19, Jan. 2022, doi: [10.1007/s00477-021-02168-w](https://doi.org/10.1007/s00477-021-02168-w).
- [113] M. E. Billah and F. Javed, "Bayesian convolutional neural network-based models for diagnosis of blood cancer," *Appl. Artif. Intell.*, vol. 2021, pp. 1–22, Dec. 2021, doi: [10.1080/08839514.2021.2011688](https://doi.org/10.1080/08839514.2021.2011688).
- [114] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, Jul. 2019, doi: [10.3390/w11071387](https://doi.org/10.3390/w11071387).



ABDULLAH A. ABDULLAH was born in Duhok, Kurdistan Region, Iraq, in 1988. He received the B.Sc. degree from the University of Duhok, Iraq, in 2009, and the M.Sc. degree from the University of Bridgeport, Bridgeport, CT, USA, in 2014. He is currently pursuing the Ph.D. degree with the University of Zakho, where he is working on probabilistic deep learning models.

From 2009 to 2012, he worked as a Programmer Assistant at the University of Zakho and as a Laboratory Mentor at the University of Bridgeport, from 2013 to 2014. Since 2014, he has been working as a Lecturer with the Computer Science Department, Faculty of Science, University of Zakho. His research interests include machine learning algorithms, probabilistic deep learning, and computer security.



MASOUD M. HASSAN 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 Lecturer at the Department of Computer Science, University of Zakho. His research interests include statistical modeling, machine learning, data mining, and artificial intelligence.



YASEEN T. MUSTAFA (Member, IEEE) received the M.Sc. degree in mathematics, with a specialization in applied mathematics, from the University of Duhok, Duhok, Kurdistan Region, Iraq, in 2005, and the Ph.D. degree in applied statistics of remote sensing from the Faculty of Geo-Information Science and Earth Observation, University of Twente (ITC), Enschede, The Netherlands, in 2012. He is currently a Professor of applied remote sensing and GIS with the

Faculty of Science, University of Zakho, Zakho, Kurdistan Region. He is also occupying the position of Vice President for scientific affairs at the University of Zakho. He supervised several projects related to remote sensing and GIS. His research interests include spatial statistical and spatiotemporal data in remote sensing and GIS, such as monitoring data, image analysis, issues of data quality, and Bayesian networks. He serves as a reviewer for a number of national and international journals. He has been involved in the editorial boards of several journals. More information can be found at: <http://yaseenmustafa.net>

...