2024 International Conference on Intelligent and Innovative Computing Applications (ICONIC)
7th-9th November 2024, Pearle Beach Resort & Spa, Flic en Flac, Mauritius
https://doi.org/10.59200/ICONIC.2024.008

Bayesian Convolutional Neural Network for Detection of Cataract Ocular Disease

Taryn Michael and Eli Nimy
Sol Plaatje University
Kimberley, South Africa
taryn.michael@spu.ac.za, eli.nimy@spu.ac.za

Seipati Nyamane and Micheal Olusanya
Sol Plaatje University
Kimberley, South Africa
seipati813@gmail.com, micheal.olusanya@spu.ac.za

Abstract
Ocular disease diagnosis using fundus images is one of the most challenging tasks in the medical field but is necessary for early screening and treatment. This manual process is extremely time consuming, complex and error prone. Currently there is an increased demand for the utilization of deep learning techniques for the automated detection of ocular diseases, especially for use on biomedical images. However, these conventional techniques such as Deep Neural Networks and Convolutional Neural Networks (CNNs) present some challenges, such as its tendency to overfit on smaller datasets and its inability to measure the uncertainty of its predictions. This is crucial in the medical field to determine the reliability of the predictions made by the automated systems. Thus, in this paper a Bayesian Convolutional Neural Network (BCNN) is implemented for Cataract disease detection to provide the reliability (uncertainty) estimates sought after in the medical field. The BCNN is benchmarked against the implementation of a standard Convolutional Neural Network. The BCNN model was compiled using the negative log-likelihood loss function and an Adam optimizer with a learning rate of 0.001 trained over 100 epochs. The CNN was compiled similarly except for the loss function which was the categorical cross entropy loss. The test results indicate that the BCNN model achieved 93.16% accuracy, while the standard CNN achieved an accuracy of 95%. Both models achieved comparable accuracy results to existing studies that utilized CNN architectures to predict ocular diseases. Although the CNN gave a slightly better accuracy, it cannot account for the uncertainty measurements of its predictions, which the BCNN is able to do. Thus, the BCNN would be more useful for ophthalmologists.

Keywords: Ocular Disease detection, Cataract, Computer Vision, Bayesian Convolutional Neural Networks.

1. Introduction

Ocular diseases refer to any type of disease or infection which negatively impacts the health and vision of a patient’s eyes (Arslan & Edras, 2023). The main ocular diseases responsible for vision impairment include cataract, glaucoma, and retinal disease (Ramanathan et al., 2021). Early detection of these diseases is crucial as early treatment can prevent it from leading to complete loss of vision or extreme damage to the eye (Leibig, et al., 2017; Arslan & Erdas, 2023). According to the World Health Organization, out of the 2.2 billion people suffering from vision impairment, at least 1 billion of them could have received early treatment to prevent it through early detection (Arslan & Erdas, 2023). The existing procedure of manual ocular disease detection is time-consuming, error-prone and complex. Thus, there is a need for an automated detection system for ocular diseases to aid in the screening process (Leibig, et al., 2017).

Currently, there is an increasing demand in the medical field to utilize deep learning and machine learning techniques for detection of diseases, including ocular diseases (Patankar, 2021; Mohammed & Farrukh, 2022; Arslan & Erdas, 2023). In particular, the use of deep learning techniques for biomedical images has been demonstrated to be a very active research area (J Goecks, et al., 2020). One of the most popular deep learning techniques for medical imaging applications, such as image segmentation, are Convolutional Neural Networks (CNNs).

CNNs removes the need for manual feature extraction and learns features directly from images (B Mohammed & J Farrukh, 2022). CNNs have already advanced to the point of surpassing human accuracy in image classification problems, thanks to its ability to fit to a wide variety of non-linear data points (Arslan & Erdas, 2023; Shridhar et al., 2019). However, CNNs require a large amount of data to train on. This thus means that standard CNNs, or any deep neural network, are highly likely to overfit on smaller datasets (E Mohammed, 2022; Shridar et al., 2019). This leads to the model fitting extremely well to the training dataset, but not performing very well on any new or test data which makes it incapable of determining any uncertainty which may be present in the training data. This causes the model to make overly confident decisions about what the correct classification or prediction may be when given new data (Shridhar et al., 2019). To reduce the model's susceptibility to overfitting, various regularization techniques are commonly employed such as L1 and L2 regularization, early stopping, and weight decay (Shridhar et al., 2019). Bayesian Convolutional Neural Networks (BCNNs), which we implement in this paper improve upon these problems and are thus a better alternative solution, especially when working with smaller datasets and in cases where uncertainty estimates are necessary.

In the medical field, it is an absolute necessity to be able to evaluate and determine the reliability of the predictions and classifications made by automated systems. Therefore, it would be beneficial for the models to be able to provide an estimate of the uncertainty of the prediction or classification made, especially since there will be some images which may be more difficult to classify due to image quality or the equipment utilized. This could enable the model to flag the images that are more difficult to classify for the medical experts to place close attention to and consult on to determine which diagnosis would be the most accurate. It would thus be unjustifiable to utilize single-point estimates for the CNN filter weights to base the classifications on (Shridhar et al., 2019).

BCNNs implement a Bayesian posterior inference over the parameters of the neural network. This essentially means that, instead of having point estimates for the filter weights or kernels, it places a probability distribution over these kernels, which enables it to be robust to overfitting. This is demonstrated in Figure 1. Through this technique, it also offers uncertainty estimates through its parameters.

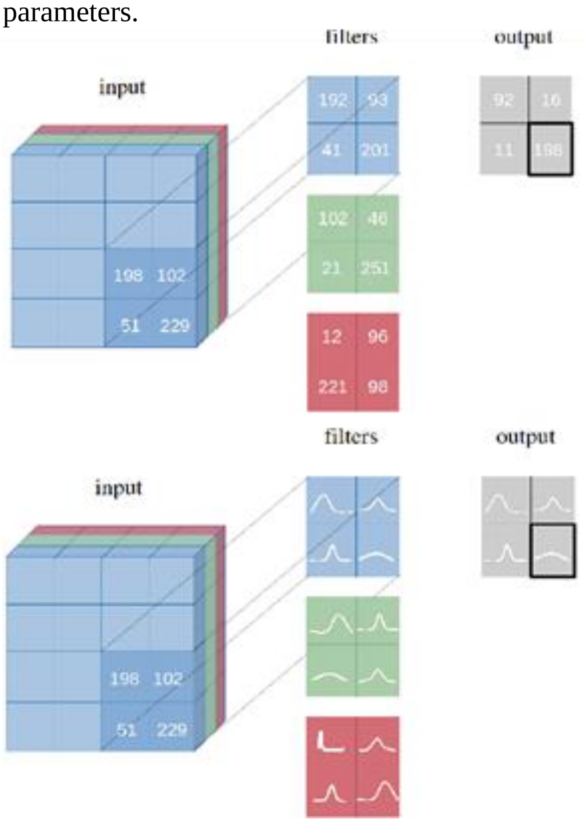

Figure 1: An example of CNN vs BCNN. The top is a standard CNN with single point-estimates as weights, and the bottom a BCNN with probability distributions over weights

In Bayesian modelling, two main types of uncertainty exist, namely Aleatoric and Epistemic. Aleatoric uncertainty is utilized to measure the uncertainty, noise, or variability present in the data collection method, for example, sensor noise. This type of uncertainty can thus not be reduced by collecting more data as it derives from the data collection method (Shridhar et al., 2019). In our case, the noise in the data could be from a dusty camera lens, or low-resolution imaging devices, and the variability could be the number of patients captured with cataracts vs no cataracts. In contrast, Epistemic uncertainty is used to measure the uncertainty of the model and can be reduced by increasing the size of the training data (Shridhar et al., 2019; Hüllermeier et al., 2021). In current literature, such as (C Leibig, et al., 2017; B Mohammed & J Farrukh, 2022), these uncertainties can be measured by placing probability distributions over the model parameters or outputs. Typically, the Epistemic uncertainty is determined through the placement of a prior distribution over the model parameters to determine to which degree these parameters or weights vary given some data. Alternatively, the Aleatoric uncertainty is determined by placing a probability distribution over the model's output (Shridhar et al., 2019).

The advantages of BCNNs and the motivation for utilizing it for this paper can be summarized as follows:

- BCNNs offer a measure of uncertainty for model predictions, which is essential in the medical field, where an incorrect diagnosis could prove fatal (C Leibig, et al., 2017).
- BCNNs tend to work well on small datasets such as the one utilized for this paper (Shridhar et al., 2019).

Thus, the development of a model which may effectively classify fundus images as having a cataract or not with a measurement of uncertainty would greatly assist in an improved cataract disease diagnosis.

The rest of the article is organized as follows: Related Works, which describes the BCNN algorithm and how existing literature is utilizing it and other computer vision methods for automated detection of ocular diseases. The methodology section provides an overview of the data and the steps followed to implement and test the BCNN. Important findings are discussed in the experiments and results section. Finally, the study draws on conclusion in section 5.

2. Related Works

A cataract is an obscuring or dulling of the lens within the eye. This is one of the most prevalent ailments that can lead to blindness. Machine Learning and Deep learning-based methods have been used to assist in mitigating severe occurrences of cataracts due to the damaging impact of cataracts. However, conventional deep learning-based methods lack the ability to provide reliable measurements of the certainties of its predictions, which is highly sought after in the medical field. Thus, BCNN was introduced. It can gain knowledge on the key features of datasets, even on smaller datasets that conventional techniques have difficulty fitting to. It can incorporate feature learning actions into the model development, reducing the incompleteness of manual structure features. It also provides uncertainty estimates, allowing one to assess the reliability of its predictions, enabling them to be used in various medical imaging paradigms (Dong, Zhang, Qiao, & Yang, 2017).

The introduction of the concept of BCNN has sparked a lot of research, particularly in the medical field, due to its added benefit of providing valuable details about uncertainty in predictions. Due to their potential to gain knowledge on meaningful attributes from data, CNNs have been demonstrated to perform better than previous state-of-the-art methods in computer vision applications (BN & Babu, 2024). However, CNNs, and neural networks in broad terms, lack uncertainty quantification and thus are easily duped by attacks. Dera et al. (2020) proposed a Bayes- Synthetic aperture radar (SAR) Net, a BCNN that can classify SAR images while considering the network's confidence in its prediction. The Bayes-SAR Net propagates the mean and covariance of the estimated likelihood function of the network parameter values given the relevant information, yielding a forecasting mean and covariance of the classifying outcome. In the presence of opposing disturbance, Bayes-SAR model achieved an accuracy that is approximately 10% higher than a normal SAR model. Chai, Bian, Liu, Li, and Xu (2021) proposed a glaucoma detection method based on Bayesian Deep Multisource Learning (BDMSL). The study aimed to improve the effectiveness of automatic diagnosis by taking uncertainty into consideration and obtaining critical details from multimodal sources of data such as medical factors, images, and texts. Multisource learning is used specifically to incorporate data from numerous sources, while Bayesian deep learning is used to acquire model uncertainty details. Based on legitimate health care data captured from one of China's best eye hospitals, the findings show that the BDMSL model outperforms other methodologies in aspects of glaucoma detection. In addition, the user study demonstrates that the BDMSL model is preferred by users (i.e., ophthalmologists).

Many researchers have been interested in the analysis of fundus images for cataract detection over the last few years (Pratap & Kokil, 2019; Qiao, Zhang, Dong, & Yang, 2017). Cataracts are one of the most common causes of blindness, accounting for more than half of all blindness. Zhang et al. (2017) used Deep Convolutional Neural Network (DCNN) to detect and rank cataract automatically. It was also used to illustrate attribute maps at the pool5 layer, highlighting their elevated empirical semantic significance and explaining the rationale behind the attribute illustration extracted by the DCNN. The proposed DCNN classification system was cross validated on up to 5620 images of population-based medical retinal fundus images gathered from hospitals. From this paper it was concluded that as the number of available samples increased, so did the DCNN classification accuracies, and the non - constant variety of levels of accuracy. This reiterates the fact that conventional deep learning techniques perform best on larger datasets. The method achieved the highest accuracy of 93.52 % in cataract detection and 86.69 % in ranking cataract respectively. In another study, instead of using images to detect cataract, Yu et al. (2019) evaluated machine learning and deep learning algorithms for automated phase classification by a manual process of segmented phases in videos of cataract surgical procedure. Four techniques were used, each with a distinct set of input data: Support Vector Machine (SVM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and CNN-RNN. Each technique was tested using a 5-fold cross-validation procedure. The overall average accuracy of the four algorithms varied from 0.915 to 0.959. While specificity was generally greater across all phases and all four algorithms ranged from 0.877 to 0.999, and precision, 0.283 to 0.963.

Yusuf, Theophilous, Adejoke, and Hassan (2019) introduced a web-based Computer Aided Diagnostic for cataract prediction system that focused on CNN that can be used by anyone outside of a hospital setting. The system design was trained on a set of data of 100 eye images obtained from Google image search findings for "regular human eyes" and "cataract in human eye". It trained a different model using the ImageNet model established in Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) using the CNN classification model. The model can now categorize eye images as "Normal" or "Cataractous". The system was intended to take images as input data and attained a Sensitivity of 69%, Specificity of 86%, Precision of 86%, F-Score of 56%, and AUC of 84.56%. The accuracy was 78%, affected by the trained model during ImageNet image classification using a deep CNN. As with (Kaur, Chetty, & Singh, 2020; Kapoor & Arora, 2022). Dense-Net and U-Net were used to detect and classify eye cataract by analyzing 200 observations of image data. The experimental feature extraction analysis outperformed single layer feature extraction and overfitting that was prevented by the propagation techniques. The proposed method achieved 89.5% and 93.3% accuracy rates, 75% and 80% sensitivity, 82% and 86% specificity for Dense-Net and U-Net, respectively.

Deep Learning has provided significant insights and advancements in cataract detection, and it continues to do so due to its high accuracy rates. It has brought benefits such as detecting cataracts before it progresses to severe stages, which can aid in early treatment and its high cataract classification rates. However, there is still a great need to investigate more effective algorithms for cataract classification. This study is significant because it experimented with a BCNN for cataract classification which has been reported to be preferred in fields such as medical due to its reliability and uncertainty measurements and offers potentially improved performance against conventional methods.

2.1 Research Gap

This research addresses uncertainty estimation for the detection of cataract ocular disease using BCNN. Conventional manual diagnostic procedures are not just time consuming but also prone to errors, which can result in inefficiencies in care. By incorporating uncertainty quantification our research aims to improve the reliability of predictions ensuring that medical resources are appropriately allocated to patients diagnosed with cataract. Additionally, this method strives to enhance the detection of cataract leading to patient outcomes and more effective utilization of healthcare resources. Despite the advancements in deep learning the absence of uncertainty estimation remains a major obstacle and our study endeavours to address this shortfall and contribute further insights, to the existing knowledge base.

3. Methodology

The method used in the study for classifying cataract and normal images of ocular disease patients consists of the following steps: 1) Data collection and understanding, 2) Data Pre-processing and Transformation, 3) Modelling and 4) evaluation. The study's experiment setup is also discussed.

a) Data Collection and Understanding

The proposed CNN and BCNN models were built using a publicly available Ocular Disease Intelligent Recognition (ODIR) dataset on Kaggle (larxel, 2020). The Dataset can be found at the following link:

https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k.

The ODIR represents a structured ophthalmic database that contains diagnostic keywords from doctors, left and right eye colour fundus photographs of 5 000 patients with their respective ages. This dataset is a representation of real-world patient information obtained by numerous medical and hospital centers in China by the Shanggong Medical Technology Co., Ltd. The Fundus images present in the dataset were captured by several market cameras such as Kowa, Canon, and Zeiss, resulting in differing image resolutions. Furthermore, 8 patient classification categories are represented in the dataset, including the normal control group (N) and 7 ocular disease classes (Cataract (C), Glaucoma (G), Diabetes (D), Age-related Macular Degeneration (A), Pathological Myopia (M), and other abnormalities or diseases (O)). The CNN and BCNN models were built to classify the cataract class due to its high prevalence in ocular disease. The distribution of patients among the 8 categories of ocular disease is show in Figure 2.

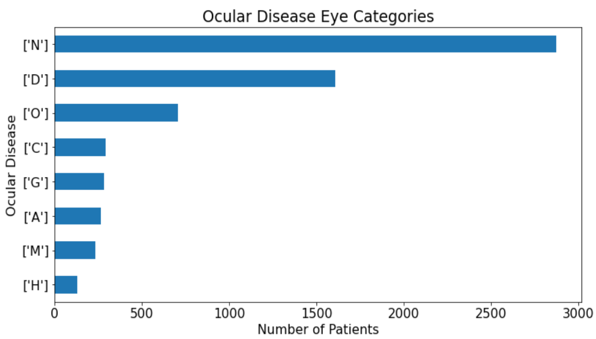

Figure 2: Patient Distribution over the 8 classification categories

b) Data Pre-processing and Transformation

Data pre-processing and transformation is an essential preliminary step taken in machine learning to convert raw input data into suitable input data for modelling (Garcia et al., 2016). The first step taken towards pre-processing was filtering out images from the classification categories of interest in the ODIR dataset, namely, the Normal and Cataract images. Since the Normal category had the most patient images, a random sample of normal images were obtained in accordance with the number of images present in the Cataract class. In total, 782 images were acquired, in which 301 images belonged to the Cataract category and 481 belonging to the Normal category. The images were randomly split into train and validation image folders in the ratio of 90:10. Out of the 782 images, 702 images were used for training and 80 images as validation data. The images were shuffled to reduce bias and variance, ensuring the BCNN models remain general and are less likely to overfit. Data loading was performed in batch sizes of 44 and 4 for training and validation, respectively.

Pre-processing and transformation steps:

1. Read Cataract and Normal image files from Train and test directory
2. Resize each image to 75×75 pixels for consistency and computational efficiency
3. Rescale each image by $\frac{1}{255}$ to ensure all the pixel values are within a similar range of [0-1] for model stability and performance

c) Modelling

Two types of CNN-based models were built, a standard CNN and a Bayesian CNN (BCNN). The standard CNN model was built to benchmark the results of the BCNN. A standard CNN was chosen as a benchmark model due its demonstration of outstanding accurate results and performance in image classification and computer vision problems in comparison to other deep neural networks (Khan et al., 2022). The BCNN and CNN models were built using the same neural network architecture for consistency and comparable results. The structure of the model architecture used is shown in Figure 3, indistinguishable to the one proposed by (Winastwan, 2020).

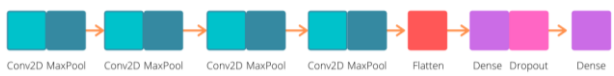

Figure 3: Standard CNN and Bayesian CNN Overall Model Architecture.

Network Architecture

The proposed standard CNN and Bayesian CNN network architecture consists of four CNN layers, one Flatten layer, a Dense and Dropout layer after flattening, followed by a Dense layer at the end. The underlining difference between the proposed Standard CNN and BCNN architecture is the first convolutional layer and last dense layer as shown in Figure 4.

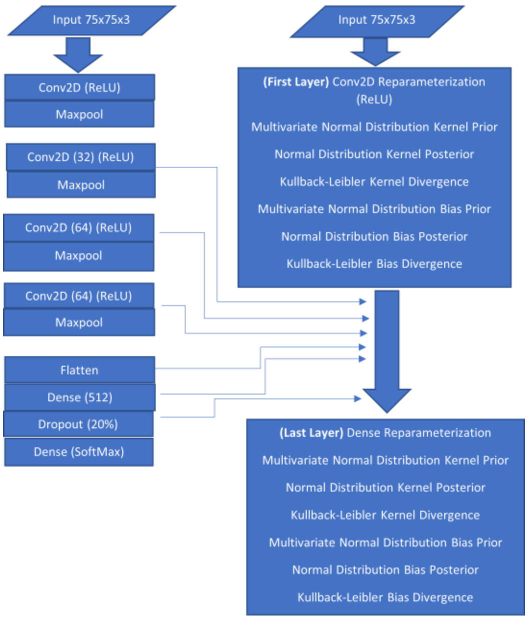

Figure 4: Standard CNN (left) and Bayesian CNN (Right) Architectures

Standard CNN

The standard CNN model was built with 4 CNN layers: 1 input CNN layer with 3 input channels and 3 hidden CNN layers. A rectifier linear unit activation function was applied to all the CNN layers. Each CNN layer output was fed to a filter size of 2×2 max pool layer. A kernel size of 3×3 dimension was used in each CNN layer with default padding and strides. The resulting output of the 2D CNN was flattened and fed to a dense layer of 512 units and a dropout layer with a rate value of 20% to avoid overfitting. Finally, a SoftMax function was used for computing the probability of each category.

ReLU Activation Function:

$$f(x) = \max(0, x)$$

The ReLU function is represented by the above equation, where x is an input to a neural network neuron. ReLU is used due to its common adoption in neural networks and its ability to overcome the gradient vanishing problem which allows models to perform better and learn faster (He et al., 2021)

SoftMax Function:

The proposed CNN model used a SoftMax Function due to the outputs being mutually exclusive. The SoftMax function was used to obtain the probability of an image belonging to a category. The SoftMax function is defined as: $f(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$

Where x_i is the input, K is the number of categories, e^{x_i} is the standard exponential function for the input, and e^{x_j} is the standard exponential of the output (Sharma et al., 2017).

Bayesian CNN

The Bayesian CNN model was built following a similar architecture as the standard CNN model with the only difference being the first convolutional layer and the last dense layer, as shown in Figure 4. A Convolutional2DReparameterization layer was used in the first layer instead of a Conv2D layer. This was to enable the BCNN model to take aleatoric uncertainty into account. This in return allows the BCNN model to create outputs drawn from a distribution in comparison to a standard CNN that creates deterministic value outputs (Winatwan, 2020). Since Bayesian models are built properly by combining probability distributions, the following was defined in the first convolutional layer (Martin, 2016): a) Prior distribution for the Kernel and bias parameters, b) Posterior distribution for kernel and bias parameters, and c) Kullback-Leibler divergence.

Prior for Kernel and bias parameters – This represents our prior belief about the kernel and bias parameters of the model before observing any data. (Martin, 2016). The prior was defined by a multivariate normal distribution that contains non-trainable parameters. The multivariate normal distribution is defined as:

$$\mathcal{N}(X|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}} exp^{-\frac{(X-\mu)^T \Sigma^{-1}(X-\mu)}{2}}$$

Where μ (mean) is the n dimensional vector, $\sum$is the covariance matrix of $n \times n$ and$|\sum|$ as the determinant of $\sum$ (Hastie et al., 2009).

Posterior for Kernel and bias parameters – This represents our updated belief about the kernel and bias parameters of the model after observing the data (Martin, 2016). The posterior was defined by a standard normal distribution that contains trainable parameters as it represents our belief after observing the data. The normal distribution is defined as: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Where σ is the standard deviation and μ is the mean (Hastie et al., 2009).

Kullback-Leibler Divergence – This was used to measure the divergence of the prior and posterior distributions (Joyce, 2011; Winastwan, 2020). In this study the $D_{KL}(P \parallel Q)$ was precisely used to measure the information gained by revising our beliefs from the prior distribution Q to the posterior distribution P. The Kullback-Leibler divergence is defined as: $D_{KL}(P \parallel Q) = \sum_{x \epsilon X} P(x) \log(\frac{P(x)}{Q(x)})$

Where P and Q are probability distributions (Joyce, 2011).

A DenseReparameterization layer was used as one of the last layers in the BCNN architecture to take epistemic uncertainty into account. Similar to the Convolutional2DReparameterization layer, the following were defined: a) Prior distribution for Kernel and bias parameters, b) Posterior distribution for kernel and bias parameters, and c) Kullback-Leibler divergence. The DenseReparameterization was then fed into a one hot categorical layer consisting of 2 units to represent the Normal and Cataract category.

Model Hyper-parameters

The BCNN model was compiled using the negative log-likelihood loss function and an Adam optimizer with a learning rate of 0.001 trained over 100 epochs. The standard CNN was compiled with the same hyper-parameters except for the loss function. A categorical cross entropy loss function was used for the standard CNN.

Evaluation

The accuracy metric was used to measure how accurately the model(s) can classify the image categories. Accuracy is defined as the number of correct classifications over the total number of classifications given by equation below (Tayal et al., 2021; Zhang et al., 2024):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP – is True Positive, TN – is True Negative, FP – is False Negative and FN – is False Negative.

Experiment Setup

Two experiments were conducted to observe the generalizability of the BCNN model to new datasets. Experiment one consisted of tuning the number neurons in each layer. In experiment two, the learning rate was tuned per BCNN model. The model checkpoint method was utilized to save models with the best validation accuracy.

4. Experiments and Results

a) Model Performance

The standard CNN model was built as a benchmark model for classification results comparison with the BCNN model using the accuracy metric. The standard CNN and BCNN models were trained over 100 epochs using an Adam optimizer with a 0.001 learning rate. The Standard CNN model achieved a 95% accuracy while the BCNN model achieved a 93.16% accuracy. Both models achieved comparable accuracy results in comparison to existing studies that utilized CNN architectures to predict ocular disease categories. The BCNN model results over 100 epochs are shown in Figure 5. The standard CNN demonstrates a better accuracy performance in correctly classifying the Normal and Cataract category. However, what is not known is how certain the standard CNN model is in correctly classifying each category. Thus, the implementation of a BCNN model to account for the uncertainties present in classifying the Normal and Cataract category.

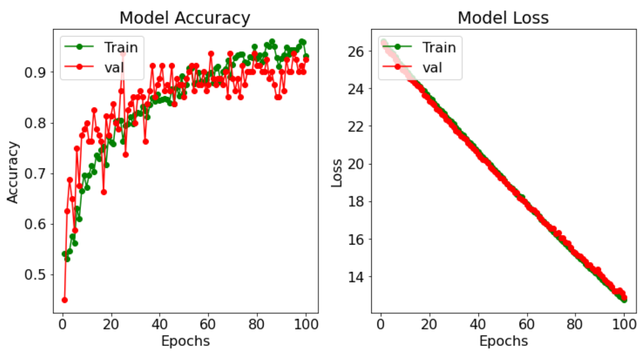

Figure 5: BCNN Model Accuracy and Loss for train and validation set over 100 epochs

b) Experiment Results

The experiments conducted in this study was to get a sense of how the BCNN model would generalize to new unseen data. Due to this, the validation accuracy was observed throughout the two experiments. Validation data was utilized in the experiments as it provides the first test against unseen data. Table 1 and 2 provides a summary of the results obtained in experiment 1 and 2.

Table 1: Experiment 1 Results-Tuning number of neurons in hidden layers

BCNN Model	Epoch Checkpoint	Validation Accuracy	Learning Rate	Number of neurons per Hidden Layer
1	37	70.00%	0.05	64
2	55	65.00%	0.05	128
3	53	72.50%	0.05	256

The experimental results in Table 1 demonstrated that a significant increase in the number of neurons per hidden layer gave a higher validation accuracy achieved at epoch 53 in comparison to 64 neurons increase in BCNN model 2.

Table 2: Experiment 2 results-Tuning learning rate

BCNN Model	Epoch Check-point	Validation Accuracy	Learning Rate	Number of neurons per Hidden Layer
1	25	93.75%	0.001	32 (1) + 64 (2)
2	85	65.00%	0.005	32 (1) + 64 (2)
3	55	63.75%	0.01	32 (1) + 64 (2)
4	87	67.50%	0.05	32 (1) + 64 (2)

The experimental results in Table 2 shows that the validation accuracy was at its highest with a 93.75% accuracy when the learning rate was 0.001. The BCNN models demonstrated better generalization with a learning rate of 0.001 and 0.05 as greater validation accuracies were achieved in comparison to the 0.01 and 0.005 learning rate used.

c) Cataract Classification: Use Case

Two cataract images from the test set are classified to demonstrate the use of a Bayesian CNN model. In Bayesian modeling, predictions are made by sampling values from the posterior distribution (Martin, 2016). Given this, each image was predicted by sampling values from the BCNN model's posterior distribution. In this study, the BCNN model classified each image 300 times, drawing 300 samples from the posterior distribution. A 97.5% prediction interval was calculated for each image, representing the range within which 97.5% of the sampled predictions fell, thereby providing a robust measure of the model's uncertainty. A narrow prediction interval indicates high certainty in the predictions, whereas a wider interval points to greater uncertainty. Additionally, the prediction intervals and the distribution of the 300 sampled predictions were visualized for each image. This visualization helps in understanding the spread and central tendency of the predictions, thereby providing a clear picture of the model's uncertainty.

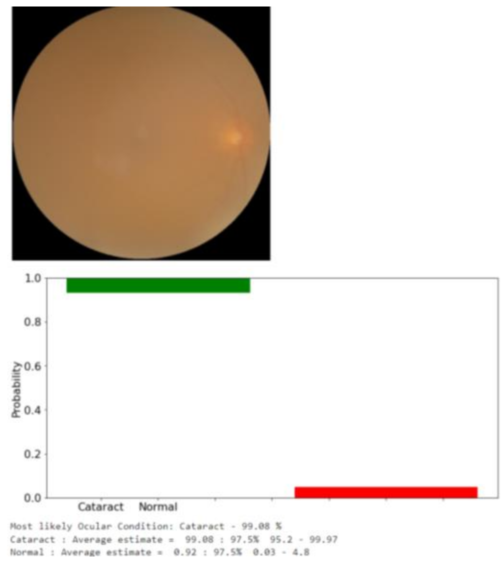

Figure 6: Cataract Image Prediction Case 1 – High Certainty

As shown in Figure 6, the BCNN model correctly predicts the Cataract category with an average percent of 99.08% with a 97.5% probability that the Cataract probability lies between 95.2% - 99.97%. The predictions were made with high certainty as the model demonstrated high certainty in assigning very high probability values to the Cataract category and very low probabilities to the Normal category as indication of the prediction outcome belonging the Cataract category. The high certainty in case 1's prediction suggests a strong confidence in the model's diagnostic capabilities. In terms of clinical decision making, predictions with high certainty can result in:

- Immediate action: clinicians can proceed with a high degree of confidence in the diagnosis of cataracts.
- Streamlined workflow: Patients with high certainty can be prioritized for treatment, which can lead to an efficient allocation of resources.

As shown in Figure 7, the BCNN model correctly predicts the Cataract category with an average percent of 54.64% with a 97.5% probability that the Cataract probability lie between 0.04% - 99.97%. The Normal category is also shown to have probability values ranging from 0.03% – 99.96%. This prediction demonstrates the existence of aleatoric and epistemic uncertainty. The aleatoric uncertainty is shown as the BCNN model also predicts the Normal category with high probability values. The epistemic uncertainty is shown as the BCNN model itself is uncertain of how big of a probability value each category should be predicted with. In terms of clinical decision making, predictions with high uncertainty can result in:

- Signal for review: high uncertainty cases should be flagged for additional review by ophthalmologists, ensuring that uncertain diagnoses are carefully evaluated.
- Request for additional testing: clinicians might order further diagnostic tests to clarify uncertain cases, reducing the risk of misdiagnosis.

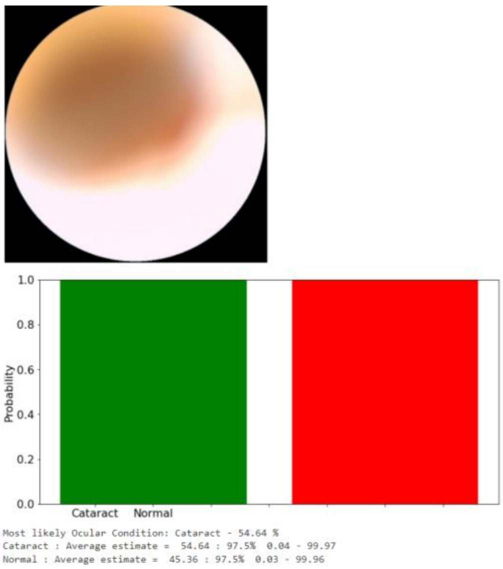

Figure 7: Cataract Image Prediction Case 2

Bayesian approaches offer significant advantages in terms of quantifying uncertainty and improving the reliability of predictions in medical image analysis. They provide valuable tools for risk assessment and help in making more informed clinical decisions. However, the computational complexity, scalability issues, and the need for careful selection of priors present notable challenges. Balancing these advantages and limitations is key to effectively leveraging Bayesian methods in medical imaging applications

6. Conclusion

Cataract is the most common eye disease in the world. If it is not detected at an early phase, it can cause blindness. Early detection and classification are the most effective ways to reduce the danger and avert painful surgery. Conventional deep learning techniques for automated disease detection have been popularly used. However, they present the challenges of overfitting on smaller datasets and being unable to provide reliability estimates of the model predictions. Thus, the purpose of this research was to investigate the usefulness of BCNN for the classification of cataract. BCNN as an automated detection method is an effective technique for cataract image classification. It aids in the goal of preventing cataracts early and improving the diagnostic efficiency of clinicians. The proposed approach was trained and tested on the Ocular Disease Intelligent Recognition (ODIR) dataset containing 5000 patient's fundus images for both eyes and doctors' diagnostic keywords. The best BCNN model produces an accuracy of 93.16 % and validation accuracy of 92.50%, which is comparable to other studies and indicates promising results BCNN has to offer in the field of ocular disease classification and detection. It is vital to improve deep learning interpretability, that further could also lead to prompt implementations in cataract analysis. More experiments can be conducted in the future to investigate the effects of image augmentation, image resizing, class imbalance handling, model checkpoint (validation loss), and the use of techniques such as transfer learning on this work. Moreover, additional research using larger datasets and encompassing various ocular diseases is essential to solidify the effectiveness of these approaches. This can lead to an increase in the accuracy of model prediction by avoiding data scarcity and creating more accurate data models with less data overfitting.

7. References

Arslan, G., & Erdaş, Ç. B. (2023). Detection of Cataract, Diabetic Retinopathy and Glaucoma Eye Diseases with Deep Learning Approach. *Intelligent Methods In Engineering Sciences*, *2*(2), 42-47.

BN, S. K., & Babu, G. S. (2024). Multiple eye disease detection using convolutional neural network. Journal of Data Acquisition and Processing, 39(1), 1310.

Chai, Y., Bian, Y., Liu, H., Li, J., & Xu, J. (2021). Glaucoma diagnosis in the Chinese context: An uncertainty information-centric Bayesian deep learning model. Information Processing & Management, 58(2), 102454.

Dera, D., Rasool, G., Bouaynaya, N.C., Eichen, A., Shanko, S., Cammerata, J., & Arnold, S. (2020). Bayes-SAR net: Robust SAR image classification with uncertainty estimation using Bayesian convolutional neural network. In 2020 IEEE International Radar Conference (RADAR) (pp. 362-367). IEEE.

Dong, Y., Zhang, Q., Qiao, Z., & Yang, J.J. (2017). Classification of cataract fundus image based on deep learning. In 2017 IEEE International Conference on Imaging Systems and Techniques (IST) (pp. 1-5). IEEE.

García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. Big Data Analytics, 1(1), 1-22.

Goecks, J., Jalili, V., Heiser, L. M., & Gray, J. W. (2020). How Machine Learning Will Transform Biomedicine. Cell, 181(1), 92–101. https://doi.org/10.1016/j.cell.2020.03.022

Hastie, T., Tibshirani, R., Friedman, J.H., & Friedman, J.H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: Springer.

He, J., Li, C., Ye, J., Qiao, Y., & Gu, L. (2021). Multi-label ocular disease classification with a dense correlation deep neural network. Biomedical Signal Processing and Control, 63, 102167.

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. Machine Learning, 110, 457–506. https://doi.org/10.1007/s10994-021-05946-3

Joyce, J.M. (2011). Kullback-leibler divergence. In International encyclopedia of statistical science (pp. 720-722). Springer, Berlin, Heidelberg.

Kapoor, P., & Arora, S. (2022). Applications of deep learning in diabetic retinopathy detection and classification: A critical review. Proceedings of Data Analytics and Management, 505-535.

Kaur, N., Chetty, G., & Singh, L. (2020). A novel approach using deep neural network vessel segmentation & retinal disease detection. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) (pp. 1-6). IEEE.

Khan, M.S., Tafshir, N., Alam, K.N., Dhruba, A.R., Khan, M.M., Albraikan, A.A., & Almalki, F.A. (2022). Deep learning for ocular disease recognition: An inner-class balance. Computational Intelligence and Neuroscience, 2022.

Khan, M.S.M., Ahmed, M., Rasel, R.Z., & Khan, M.M. (2021). Cataract detection using convolutional neural network with VGG-19 model. In 2021 IEEE World AI IoT Congress (AIIoT) (pp. 0209-0212). IEEE.

Larxel. (2020). Ocular Disease Recognition. Kaggle. Available at: https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k [Accessed: 9 April 2024].

Leibig, C., Allken, V., Ayhan, M.S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. Scientific Reports, 7. https://doi.org/10.1038/s41598-017-17876-z.

Martin, O. (2016). Bayesian analysis with Python. Packt Publishing Ltd.

Mohammad Ehtasham Billah, & Farrukh Javed. (2022). Bayesian convolutional neural network-based models for diagnosis of blood cancer. Applied Artificial Intelligence, 36(1). https://doi.org/10.1080/08839514.2021.2011688

Patankar, A. (2021). Diagnosis of ophthalmic diseases in fundus image by using deep learning. International Journal for Research in Applied Science and Engineering Technology, 9, 1842-1847.

Pratap, T., & Kokil, P. (2019). Automatic cataract detection in fundus retinal images using singular value decomposition. In 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET) (pp. 373-377). IEEE.

Pratap, T., & Kokil, P. (2019). Computer-aided diagnosis of cataract using deep transfer learning. Biomedical Signal Processing and Control, 53, 101533.

Qiao, Z., Zhang, Q., Dong, Y., & Yang, J.J. (2017). Application of SVM based on genetic algorithm in classification of cataract fundus images. In 2017 IEEE International Conference on Imaging Systems and Techniques (IST) (pp. 1-5). IEEE.

Ramanathan, G., Chakrabarti, D., Patil, A., Rishipathak, S., & Kharche, S. (2021). Eye disease detection using machine learning. In 2nd Global Conference for Advancement in Technology (GCAT) (pp. 1-5). IEEE. https://doi.org/10.1109/GCAT52182.2021.9587740.

Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. Towards Data Science, 6(12), 310-316.

Shridhar, K., Laumann, F., & Liwicki, M. (2019). A comprehensive guide to Bayesian convolutional neural network with variational inference. arXiv preprint arXiv:1901.02731.

Tayal, A., Gupta, J., Solanki, A., Bisht, K., Nayyar, A., & Masud, M. (2021). DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases. Multimedia Systems, 1-22.

Winastwan, R. (2020). Creating image classification model with Bayesian perspective. Towards Data Science. Available at: https://towardsdatascience.com/creating-image-classification-model-with-bayesian-perspective-a90a5956b14e [Accessed: 5 April 2024].

Yadav, J.K.P.S., & Yadav, S. (2022). Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning. Computational Intelligence.

Yu, F., Croso, G.S., Kim, T.S., Song, Z., Parker, F., Hager, G.D., Reiter, A., Vedula, S.S., Ali, H., & Sikder, S. (2019). Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. JAMA Network Open, 2(4), e191860-e191860.

Zhang, L., Li, J., Han, H., Liu, B., Yang, J., & Wang, Q. (2017). Automatic cataract detection and grading using deep convolutional neural network. In 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC) (pp. 60-65). IEEE.

Zhang, X., Wu, X., Xiao, Z., Hu, L., Qiu, Z., Sun, Q., & Liu, J. (2024). Mixed-decomposed convolutional network: A lightweight yet efficient convolutional neural network for ocular disease recognition. CAAI Transactions on Intelligence Technology, 9(2), 319-332.