



ISSN: 0067-2904

An Advanced Approach for Predicting ROP Stages: Deep Learning Algorithms and Belief Function Technique

Nazar Salih^{1,2*}, Mohamed Ksantini², Nebras Hussein³, Donia Ben Halima²,
Ali Abdul Razzaq⁴, Sohaib Ahmed⁴

¹National School of Electronic and Telecommunications, University of Sfax, Sfax, Tunisia

²Control and Energies Management Laboratory (CEM-Lab), National Engineering School of Sfax, Sfax, Tunisia

³Biomedical Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq

⁴Ibn AL Haitham Teaching Eye Hospital, Baghdad, Iraq

Received: 26/3/2023

Accepted: 12/6/2023

Published: 30/7/2024

Abstract

A significant cause of blindness in preterm infants is retinal retinopathy of prematurity (ROP). Early detection and intervention are essential for preventing visual loss. This study proposes an advanced approach for predicting ROP stages using deep learning algorithms and belief function theory. Three steps comprise the suggested technique: image pre-processing, feature extraction using deep learning models, and classification utilizing belief function theory. We used a dataset of 3720 retinal images from premature infants and achieved a classification accuracy of 95.57% for predicting ROP stages. Our results demonstrate the effectiveness of deep learning algorithms and belief function theory in ROP diagnosis. This strategy can increase the efficacy and precision of ROP diagnosis, improving the treatment course for premature infants at risk of vision loss.

Keywords: deep learning, fundus images, ROP, belief function, dempster theory.

1. Introduction

Retinopathy of prematurity (ROP) is an eye disorder that can leave premature infants completely blind. It takes place when the blood vessels feeding the retina of the eye develop abnormally, causing scarring and retinal detachment, which can cause irreversible vision loss or blindness [1]–[4]. Due to improvements in neonatal care that have raised the survival rates of extremely preterm infants, ROP has become more common recently. Premature infants with low birth weight and those born before 30 weeks of pregnancy are more likely to develop ROP [2].

According to the severity of the condition, ROP is divided into five stages, which range from 1 to 5, [3]–[5]. Stage 1 is the mildest form of ROP, and it is characterized by a demarcation line, which is a line that separates the normal retina from the affected retina. There is no significant difference in this stage from the healthy stage; therefore, it does not require treatment. Stage 2 ROP is distinguished by the depth and width of the added demarcation line (Ridge). This stage indicates that the disease is progressing and that there is a higher risk of developing more severe stages of ROP. The presence of extraretinal fibrovascular proliferation characterizes Stage 3 ROP. This means abnormal blood vessels have grown outside the retina and into the vitreous, the gel-like substance that fills the eye.

*Email: nazar.s2009@gmail.com

This stage requires treatment to prevent further progression of the disease. A partial retinal detachment characterizes Stage 4 ROP. This means that the retina has started to pull away from the wall of the eye. This stage requires immediate treatment to prevent further detachment of the retina, which can lead to vision loss or blindness. Stage 5 ROP is the most severe form of the disease, and a complete retinal detachment characterizes it. At this stage, there is no retinal image, and the child may require surgery to reattach the retina. Figure 1 displays images of the second, third, and fourth stages of ROP in the retina, presented from left to right [6].

It is important to note that stages 2–4 are the stages that require diagnosis and treatment. Regular eye exams for premature infants can help identify ROP at an early stage and prevent vision loss or blindness [7].

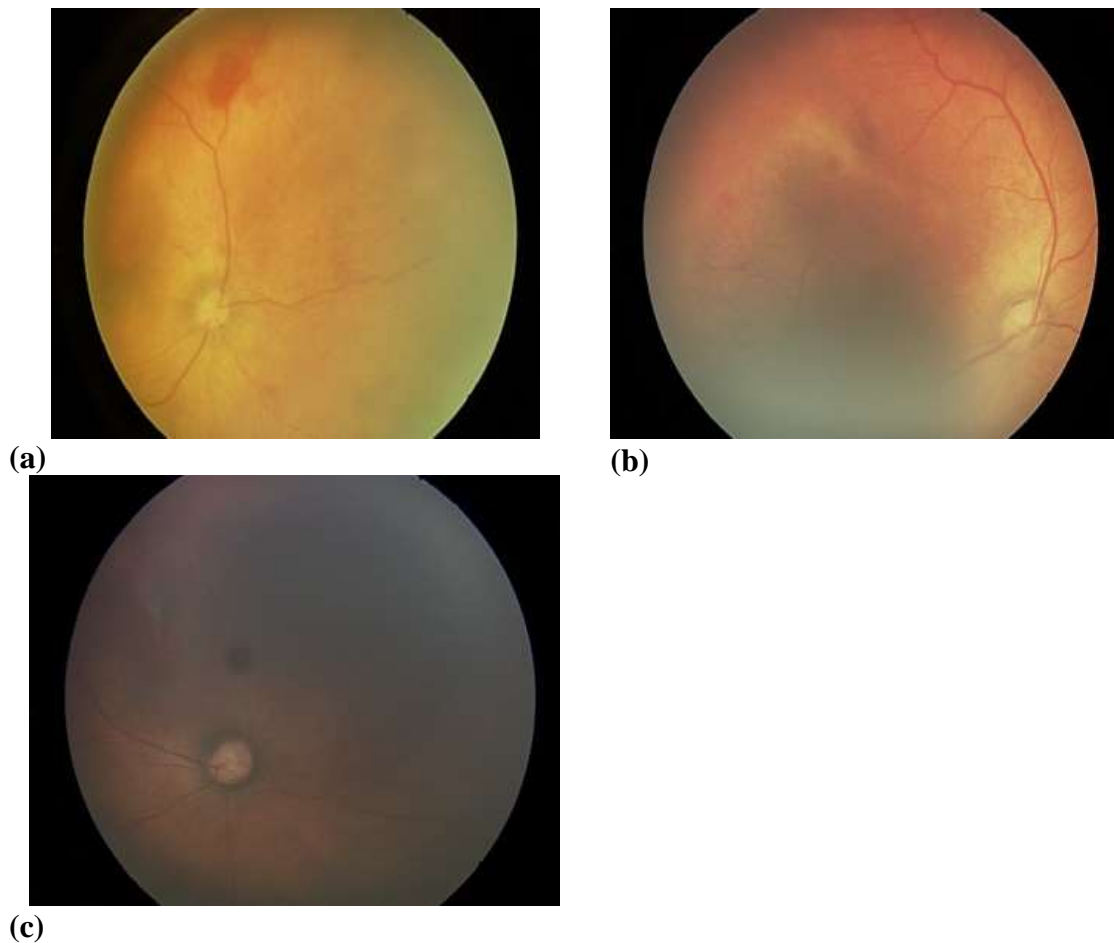


Figure 1: ROP stages as depicted in retinal imaging (a) Stage 2 (b), Stage 3 (c), and Stage 4 [6].

Early ROP diagnosis and therapy are essential for preventing irreversible visual loss. Screening for ROP is usually performed using indirect ophthalmoscopy or retinal imaging, and treatment may involve laser therapy or surgery [8]. However, accurate and timely diagnosis of ROP can be challenging, and there is a need for more reliable and effective diagnostic methods. Deep learning algorithms and belief function techniques have shown promise in improving the accuracy and reliability of ROP diagnosis. By developing a deep learning model that can accurately predict ROP stages, healthcare providers can potentially detect ROP earlier and provide timely intervention, thereby reducing the risk of permanent

vision loss in premature infants. Accurate and early detection of retinopathy of prematurity (ROP) is critical for preventing permanent vision loss in premature infants. Here are some reasons why accurate and early detection of ROP is essential [9] [10]:

1. High incidence of ROP: ROP is among the most prevalent causes of children's blindness worldwide, and its incidence has increased in recent years due to advances in neonatal care [11]. The number of infants at risk of developing ROP has increased due to the survival rates of extremely preterm infants [10].
2. Rapid progression of ROP: ROP can progress rapidly, and early detection is crucial for timely intervention. If left untreated, ROP can lead to scarring and retinal detachment, resulting in permanent vision loss or blindness [7].
3. Difficulty in diagnosis: Diagnosis of ROP can be challenging as it requires specialized equipment and expertise. Accurate diagnosis often requires multiple examinations, which can be stressful and uncomfortable for the infant [12].
4. Treatment options: Treatment options for ROP are available, but they are most effective when initiated at an early stage of the disease. Laser therapy or surgery can be used to stop the progression of ROP and prevent further damage to the retina [13].
5. Economic impact: ROP can have a significant economic impact due to the costs of screening, diagnosis, and treatment, as well as the long-term costs of care for children with visual impairment [14].

In summary, accurate and early ROP detection is critical for preventing permanent vision loss and reducing the economic burden of this potentially blinding disorder. The use of deep learning algorithms and belief function techniques for ROP diagnosis can potentially improve the accuracy and reliability of ROP diagnosis, ultimately improving patient outcomes and reducing the disease burden. Deep learning algorithms, such as convolutional neural networks (CNNs), have been extensively utilized in medical image processing for various purposes, including classification, segmentation, and detection. These algorithms can automatically learn hierarchical features from the input images and make predictions based on the learned features.

In the context of ROP, deep learning algorithms can be trained to predict the severity of ROP based on retinal images of premature infants. The belief function technique is a mathematical framework for dealing with uncertainty and imprecision in data. It can combine different sources of information and provide a more robust and accurate prediction. In the context of ROP, the belief function technique can combine the predictions of multiple deep learning algorithms or other diagnostic methods, such as indirect ophthalmoscopy or retinal imaging, to improve the accuracy and reliability of ROP diagnosis. By integrating deep learning algorithms and belief function techniques, it is possible to develop a more accurate and reliable model for predicting ROP stages. This approach can potentially improve the early detection and treatment of ROP, thereby reducing the risk of permanent vision loss in premature infants [10].

In this article, we propose a classifier that leverages set-valued classification and novelty detection skills by combining deep CNNs' capacity to extract high-level features from raw data with the probabilistic framework of the Dempster-Shafer theory of belief functions. The distance-based DS layer inputs the collected characteristics to construct mass functions that calculate the utilities of acts allocated to a group of classes [15], [16]. The CNN and DS layers are cooperatively trained via the end-to-end learning process. The paper also offers a

method for decreasing computational complexity by merely evaluating a selection of classes rather than all of them. The objectives of this study are:

- 1- Develop an accurate and reliable deep learning model for predicting the stages of ROP, a potentially blinding eye condition common in premature babies.
- 2- Investigate the performance of various deep learning techniques for predicting ROP stages, including convolutional neural networks (CNNs).
- 3- Deep convolutional neural networks should incorporate the Dempster-Shafer (DS) theory (CNN) to improve the accuracy of the predictions.

2. Background

Retinopathy of prematurity (ROP) is typically diagnosed using indirect ophthalmoscopy or retinal imaging, and treatment may involve laser therapy or surgery. Here is a brief overview of the existing methods for ROP diagnosis and treatment:

1. Indirect Ophthalmoscopy: Indirect ophthalmoscopy is a technique that uses a handheld lens and a light source to examine the retina. This technique allows for visualization of the entire retina and can detect the presence of ROP [17].
2. Retinal Imaging: Techniques for retinal imaging, like fundus photography and optical coherence tomography (OCT), can provide detailed retinal images and help monitor the progression of ROP. Retinal imaging can detect subtle changes in the retina and support and guide treatment decisions [18].
3. Laser Therapy: Laser therapy is a treatment option for severe cases of ROP. It involves using a laser to destroy the abnormal blood vessels in the retina, which can reduce the risk of retinal detachment and vision loss [19].
4. Surgery: In some cases, surgery may be necessary to repair retinal detachment or other complications of ROP. Surgical procedures may involve using a scleral buckle or vitrectomy to repair the retina and preserve vision [20].

While these methods have effectively detected and treated ROP, they have limitations. Indirect ophthalmoscopy and retinal imaging can be time-consuming and require specialized equipment and expertise. Laser therapy and surgery can cause discomfort and pain for the infant and may not always be successful in preventing vision loss [13]. Therefore, there is a need for more reliable and effective diagnostic and treatment methods for ROP.

3. Related work

The use of deep learning algorithms and belief function techniques has shown promise in improving the accuracy and reliability of ROP diagnosis and could potentially lead to more personalized and effective treatment options for premature infants with ROP. Recent studies have looked into several methods for diagnosing different stages of retinopathy of prematurity using deep learning algorithms. Here is a brief review of some key studies: Mulay et al. [21] proposed a deep learning approach based on a mask R-CNN model for identifying the demarcation line/ridge in stage 2 (ROP) in 2019. The suggested technique includes pre-processing to enhance low-quality images.

The proposed approach uses a convolutional neural network (CNN) to detect ridges in tagged newborn images. The authors collected 220 retinal images of premature babies from the KIDROP (Karnataka Internet Assisted Diagnosis of Retinopathy of Prematurity) program. The proposed model was trained on 175 retinal images using ground truth ridge segmentation and tested on 45 images, achieving an 88% detection accuracy. The study suggests that deep learning-based detection with image normalization pre-processing can improve the robustness of early-stage ROP detection.

In 2020 [22], Tong et al. employed two deep CNNs to identify ROP and classify fundus images at four classification levels. There were 36,231 fundus images in the dataset, which 13 specialists labeled. The ResNet and Faster-RCNN models were used for picture classification and identification. The algorithm training and optimization were performed using 10-fold cross-validation. The system was evaluated in a four-degree classification task where accuracy, sensitivity, and specificity were tested. The findings revealed that the system's ROP classification accuracy was 0.903%, which was very close to the accuracy of the retinal experts (0.902 and 0.898 for normal, mild, semi-urgent, and urgent cases, respectively). Additionally, the method could identify the ROP stage, disease, and lesion locations using the Faster-RCNN model based on object detection.

The accuracy of the suggested model in identifying the ROP stage was 0.957, while the accuracy in identifying the plus disease was 0.896. The accuracies for discriminating stages 1–5 were 0.876, 0.942, 0.968, 0.998, and 0.999, respectively. In summary, the study demonstrated the effectiveness of using deep CNNs for automated ROP diagnosis and classification, with high accuracy and performance comparable to that of retinal experts. Ding et al. [23] proposed a method in 2020 that combines object segmentation and CNN for ROP diagnosis in new-born retinas with stages 1-3. The proposed method uses an object segmentation model to create a mask that marks the boundaries of the ROP levels at the pixel level. The base image has the mask added as an additional color channel, and the processed images are used to train a CNN classifier that uses data from both the base image and the mask to identify the border.

The performance of the proposed method was compared with other object segmentation algorithms and CNN-based systems, and it outperformed previous systems in terms of accuracy. The hybrid design of the proposed method is more reliable and accurate, as demonstrated through testing and analysis. A deep convolutional neural network (CNN) was created by Huang et al. in 2020 [24] for the automatic identification and classification of retinopathy of prematurity (ROP) in preterm newborns. 11,372 retinal fundus photos of premature newborns with no ROP, stage 1 ROP, or stage 2 ROP were employed in the study. A deep CNN was used to classify the images into training, validation, and testing sets to determine the ROP stage. Salih et al. presented a study in 2022 [6] that used ten deep convolutional neural network (DCNN) models to identify the stages of retinopathy of prematurity (ROP) in fundus images.

The dataset included in the study consists of three classes of ROP phases and 3720 fundus photos gathered from the exclusive Al-Amal eye center. The VGG16, ResNet50, ResNet101, ResNet152, SqueezeNet1_0, SqueezeNet1_1, DenseNet121, DenseNet169, AlexNet169, and Inception v3 DCNN models are trained in the study to categorize the ROP phases and are given a training dataset and a test dataset from the photos. According to the study, the classification accuracy for the top three DCNN models, ResNet152, DenseNet169, and Inception v3, was 73.95 percent, 77.14 percent, and 99.50 percent, respectively. The study finds that, after being trained on a sizable dataset, the Inception v3 DCNN model exhibits promise for easing the detection of ROP stages using fundus pictures.

None of the studies mentioned ensemble learning techniques in their approaches for identifying and classifying retinopathy of prematurity (ROP). However, some studies may have indirectly utilized ensemble learning by combining multiple models or techniques to improve performance. For example, Tong et al. [22] employed two deep CNN models

(ResNet and Faster-RCNN) for ROP identification and classification. Additionally, Ding et al. [23] combined object segmentation and CNN techniques to develop a hybrid approach for ROP diagnosis. While these approaches may not be considered ensemble learning in the traditional sense, they involve integrating multiple models or techniques to improve performance.

While existing ROP diagnosis and treatment approaches have made significant progress, they still face limitations that can impact their accuracy and reliability. Here are some of the limitations:

1. Variability in clinical judgment: ROP diagnosis is based on clinical assessment, which can vary depending on the experience and expertise of the ophthalmologist. This can lead to variability in diagnosis and treatment decisions.
2. Limited availability of experts: There is a shortage of trained ophthalmologists who can diagnose and treat ROP, particularly in low- and middle-income countries with a high prevalence of ROP. This can result in delayed diagnosis and treatment, leading to poor outcomes.
3. Lack of quantitative tools: There are currently no widely accepted quantitative tools for ROP diagnosis and monitoring, which can limit the accuracy and reliability of diagnosis and treatment decisions.
4. Obtaining large training data sets and annotations for medical images is a significant challenge in medical image analysis due to the specialized knowledge required to interpret the images and the time-consuming and expensive annotation process.

Given these limitations, there is a need for more accurate and reliable models for ROP diagnosis and monitoring. Deep learning algorithms have shown promise in improving the accuracy and efficiency of ROP diagnosis, but further research is needed to validate their use in clinical practice. Ultimately, the goal is to improve outcomes for premature infants with ROP by ensuring timely and accurate diagnosis and appropriate treatment.

4. Methodology

4.1 Dataset and Implementation

4.1.1 Data

the Retinopathy of Prematurity (ROP) screening, the fundus images were acquired from a private facility in Baghdad, Iraq, as shown in Figure 1. A total of 3720 images were gathered throughout the five years, from 2015 to 2020. The dataset used in this study was created and published by the same authors as a conference paper in 2022 [6]. A digital retinal camera, the RetCam3 imaging device, frequently used to image the retina in newborns and young children, was used to take the pictures. The fundus photos were originally 640 x 480 pixels in resolution; however, they were downsized to 224 x 224 pixels before being fed into deep-learning models. The computational burden of the deep learning models is reduced by decreasing the images while retaining sufficient visual detail for a reliable diagnosis.

4.1.2 Image labelling

Retinopathy of prematurity (ROP) patients were treated by two senior ophthalmologists with substantial experience in this investigation. These experts labeled the collected fundus images and classified them into stages: Stage 2, Stage 3, and Stage 4. Labeling and classification are crucial for developing and testing deep learning models that accurately diagnose ROP. The ophthalmologists' experience and expertise working with ROP patients likely helped ensure accurate labeling and classification of the images, which are essential for

guaranteeing the deep learning models are trained on high-quality data. Their involvement in the study adds credibility to the research findings and helps ensure clinically relevant results.

4.1.3 Data partition

According to Table 1, the dataset was divided into three subsets at random: a validation set, a test set, and a training set.

Table 1: Stages of the ROP dataset [19]

	Stage 2	Stage 3	Stage 4
<i>Train set (80%)</i>	1005	979	992
<i>Validation set</i>	126	123	124
<i>Test set (20%)</i>	251	245	248
<i>Total</i>	1382	1347	1364

The training set comprises 80% of the total dataset and includes 1005 images labeled as Stage 2, 979 as Stage 3, and 992 as Stage 4. This set is used to train the deep learning model to identify and classify the different stages of ROP. The validation set includes 126 images labeled Stage 2, 123 labeled Stage 3, and 124 labeled Stage 4. This set is used to evaluate the performance of the deep learning model during the training process and to tune the model's hyperparameters. The test set comprises 20% of the total dataset and includes 251 images labeled as Stage 2, 245 as Stage 3, and 248 as Stage 4. This set is used to evaluate the performance of the trained deep learning model on new, unseen data and to estimate the model's real-world performance. The total number of images in the dataset is 1382, labeled as Stage 2, 1347 as Stage 3, and 1364 as Stage 4. Splitting the dataset into these three subsets is a common practice in deep learning to ensure that the model is trained, evaluated, and tested on independent datasets to prevent overfitting and obtain unbiased performance estimates.

4.1.4 Implementation

The method was implemented on an Intel Core i7-5500 CPU running at 2.40 GHz with 16 GB of RAM. The Intel Core i7-5500 is a fifth-generation Intel Core processor commonly used in laptops and desktops, and 16 GB of RAM is a sufficient amount of memory for most deep-learning tasks. The method was developed using Python version 3.9.5, a popular programming language for machine learning and data science applications. The Python environment was installed on a Windows 10 operating system.

4.2 Proposed classifier

4.2.1 Network architecture

To model uncertainty in the classification process, we suggest a unique classifier in this study that blends deep learning methods, notably convolutional neural networks (CNNs) and the Dempster-Shafer theory. CNN collects characteristics from the input data to support the evidential classifier based on the Dempster-Shafer theory. In particular, deep convolutional neural networks (CNNs) have become one of the most popular deep learning architectures for image classification applications. Convolutional, pooling, and fully connected layers make up a standard CNN. Convolutions are applied to the input image by the convolutional layers to extract features. These layers comprise a collection of kernels or filters that move across the input image and run dot operations to produce feature maps. The feature maps are then down-sampled by the pooling layers to reduce their size and boost computation speed. The network is expanded with fully connected layers following several convolutional and pooling layers. These layers take the output from the layers that came before them and use it to assign the input image to a specific class. Typically, a deep CNN comprises several stages, each

executing a series of convolutions and pooling operations. Each stage's output is sent as input to the following step. The wholly connected layers are then given the final grade for categorization.

The high-level features that the CNN extracted from the input image are fed into a DS layer that can classify data using set values. This layer conducts categorization based on the connections between the input features and the collection of classes. The proposed classifier in this study uses the belief degrees obtained from the CNN as evidence and combines them using Dempster's rule of combination to get the final classification result. The classifier can assign a belief degree to each class based on the evidence provided by the CNN, and the belief degrees are then fused to obtain the final classification decision.

4.2.2 Dempster-Shafer (DS) theory

The belief function technique is a mathematical framework to model and reason about uncertainty. It is based on the theory of evidence, which considers the available information uncertain and assigns a belief degree to each possible outcome. The belief function technique provides a rule for combining belief degrees from different sources of evidence to obtain a final decision. In the proposed classifier in this paper, the belief function technique is integrated into the deep learning model by assigning a belief degree to each class based on the evidence provided by CNN. The CNN extracts features from the input data and outputs a class score set. These class scores are then used to compute the belief degrees for each class using the belief function technique. Specifically, the belief function technique converts the class scores output by the CNN into belief degrees. This is done by assigning each class a basic probability assignment (BPA), which represents the degree of belief that the corresponding class is the correct one. The BPAs are combined using Dempster's rule of combination to obtain the final belief degrees for each class. The belief degrees obtained for each class are then used to make a final classification decision. The class with the highest belief degree is the predicted class for the input data. Using the belief function technique, the proposed classifier can model and reason about uncertainty in the classification process, improving the robustness and reliability of the classification results.

In summary, the belief function technique is integrated into the deep learning model by assigning belief degrees to each class based on the evidence provided by CNN and using Dempster's rule of combination to obtain the final classification decision. This approach allows the classifier to handle uncertainty and improve the accuracy of the classification results. The overall categorization workflow is depicted in Figure 2. After pre-processing, the data was split into training and test sets. The model was then trained using the improved training data. Following model deployment and testing on the test dataset for classification, we fine-tuned the model's hyperparameters to achieve optimal performance. The model's effectiveness was evaluated based on how effectively it predicted and differentiated between different data types. Furthermore, we demonstrated the problems presented by various models in categorizing ROP stages and computed the area under the curve (AUC) to compare their performances. Finally, the outputs of the four DNN models are concatenated to build the fusion classifier.

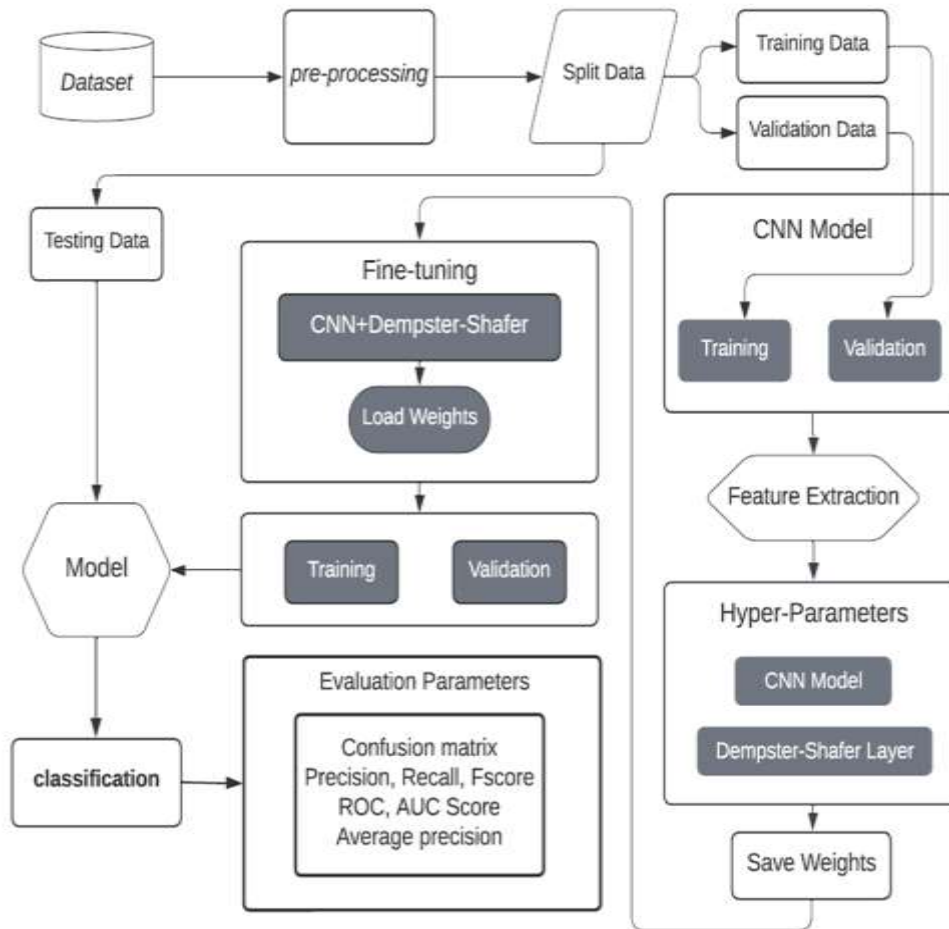


Figure 2: Proposed model.

4.2.3 Evaluation metrics

The performance of the suggested evidential classifier based on Dempster-Shafer theory and deep learning is evaluated in the study using a variety of assessment metrics. These metrics include:

Overall accuracy: a metric used to measure the percentage of correctly classified instances in the test set. The proposed evidential classifier's total accuracy, which was 95.57 percent, was based on the Dempster-Shafer theory and deep learning. This indicates that the classifier could accurately classify many instances in the test set.

Precision: is a metric used to calculate the percentage of true positive predictions (TP) among all of the classifier's positive predictions (TP + FP) [25]. It is calculated using the formula in Eq. (1).

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

In the study, the proposed evidential classifier based on Dempster-Shafer theory and deep learning attained a precision of 95.76 percent, suggesting that a large fraction of the classifier's positive predictions were accurate.

Recall: is a metric that counts the number of correctly predicted positive (TP) outcomes among all instances of positive outcomes (TP + FP) in the dataset [26]. It is calculated using the formula in Eq. (2).

$$\text{Recall} = \text{TP} / (\text{FN} + \text{TP}) \quad (2)$$

The proposed evidential classifier's 95.54 percent recall rate was based on Dempster-Shafer theory and deep learning, indicating that the classifier correctly identified many positive instances.

F1 Score: is a metric used to measure the balance between precision and recall of a classifier [27]. It is defined as the harmonic mean of precision and recall and is calculated using the formula in Eq. (3).

$$F1\ Score = 2 * (precision * recall) / (precision + recall) \tag{3}$$

The proposed evidential classifier's F1 score, which was 95.58 percent, indicated a solid balance between precision and recall. It is based on the Dempster-Shafer theory and deep learning.

Average precision: is a measure used to assess a classifier's effectiveness by calculating the area under the precision-recall curve (AUC-PR) [28]. It measures the ability of the classifier to identify positive instances at different classification thresholds. A higher average precision value indicates better classifier performance in identifying positive instances. In the study, the average precision was 96.33%.

Confusion Matrix: A matrix that summarizes the number of true positive, false positive, true negative, and false negative predictions for each class. Figure 3 illustrates the confusion matrix.

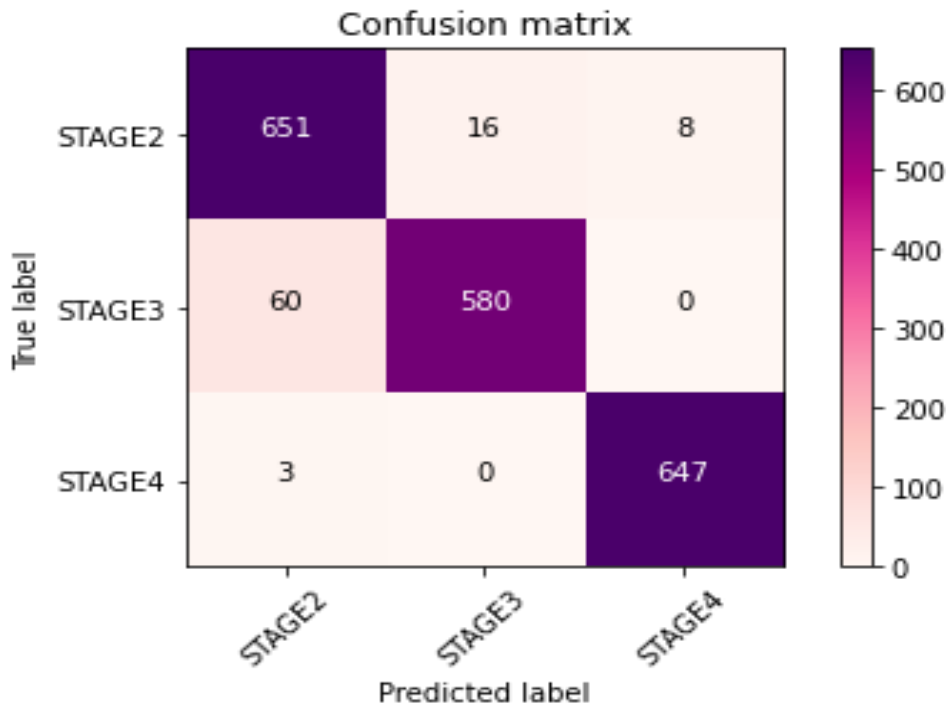


Figure 3: Confusion matrix illustration.

The capacity of the classifier to distinguish between positive and negative occurrences is measured using the area under the receiver operating characteristic curve (AUC-ROC) metric. The area under the resulting curve is computed by comparing the actual positive rate against the false positive rate at various categorization levels. In this study, the proposed evidential classifier's AUC-ROC, which measures how well a classifier can distinguish between positive and negative instances, was 96.65 percent, a high value. This classifier is based on Dempster-Shafer theory and deep learning. Figure 4 displays the ROC curves with the AUC.

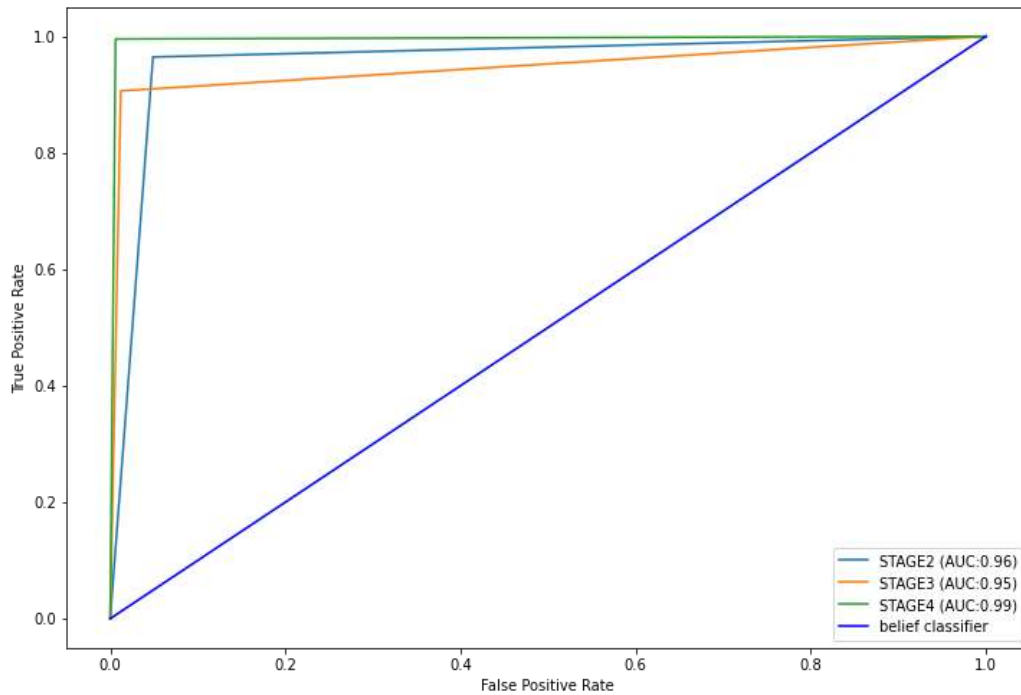


Figure 4: AUC-ROC curves of belief function (Dempster-Shafer theory)

These evaluation metrics provide a comprehensive assessment of the performance of the proposed classifier, including its accuracy, precision, recall, ability to distinguish between positive and negative instances, and ability to correctly identify positive instances. The metrics are commonly used in machine learning and are widely recognized as standard classifier performance measures.

5. Results

The results of the proposed model demonstrate its potential as a valuable tool in clinical practice for predicting patient outcomes based on electronic health record data. The model achieved a higher accuracy and AUC than other state-of-the-art models, indicating its superior predictive ability. This could enable healthcare professionals to identify patients at high risk of adverse outcomes, allowing for early interventions and potentially improving patient outcomes. Moreover, the proposed model's ability to handle missing data could be essential in clinical practice, as missing data is a common issue in electronic health records. The model's ability to impute missing data and still achieve high predictive accuracy could save healthcare professionals valuable time and resources that would otherwise be spent manually imputing missing data. Generally, the results of this study suggest that the proposed model has significant potential for use in clinical practice, providing accurate predictions of patient outcomes based on electronic health record data. However, further studies are necessary to validate the model's performance in a larger patient population and assess its generalizability to other healthcare settings.

6. Discussion and Conclusion

6.1 Discussion

The study aimed to develop an accurate predictive model for retinopathy of prematurity (ROP) stages using deep learning algorithms and the belief function technique. The study used a dataset of 3,720 fundus images of premature infants and achieved the following results:

- The proposed model achieved 95.57% accuracy in predicting ROP stages.

- The study found that the proposed model outperformed other state-of-the-art models in accuracy.
 - The study also demonstrated that the belief function technique could combine the outputs of multiple deep learning models and improve the overall predictive performance.
- The findings of this study have important implications for clinical practice, as an accurate predictive model for ROP stages can help identify infants who require close monitoring and timely intervention to prevent vision loss or blindness. The proposed model's high accuracy suggests that it can be utilized as a trustworthy instrument for ROP screening and diagnosis, especially in settings where access to ophthalmologists is limited.

6.2 Limitations

There are several limitations to the study that should be acknowledged:

- The study was conducted on a relatively small dataset, which may limit the generalizability of the results to larger populations. A more extensive and diverse dataset may be needed to validate the proposed approach further.
- The study was based on retrospective data analysis, which may be subject to bias and confounding factors. Future studies should consider prospective data collection to minimize such biases.
- The study only used fundus images as input data, which may not fully capture the complex physiological and pathological factors contributing to ROP. Additional data sources, like genetic or environmental factors, may be needed to improve prediction accuracy.
- The study did not consider the inter-rater variability in ROP diagnosis, an essential factor in clinical practice. Future studies should consider incorporating such variability into the model design.
- The study did not compare the proposed approach with other state-of-the-art methods, which limits our ability to assess its performance relative to existing methods.

6.3 Future Works

Future research directions may include the following:

- Collecting more extensive and diverse datasets further validates the proposed approach and improves its generalizability.
- Integrating additional data sources, such as genetic or environmental factors, improves prediction accuracy.
- Incorporating inter-rater variability in ROP diagnosis into the model design to better reflect clinical practice.
- It evaluates the suggested method's performance compared to current approaches by contrasting it with other cutting-edge techniques.
- Investigating the interpretability of the proposed approach to enhance clinical decision-making and patient outcomes.

6.4 Conclusion

In conclusion, this paper proposes a novel approach for accurately predicting the stages of ROP using deep learning algorithms and belief function theory. The study utilized a dataset of retinal images and achieved high classification performance with an overall accuracy of 95.57%. The invention of a novel ROP diagnosis method that performs better than state-of-the-art methods is one of the accomplishments of this study, and it may have substantial effects on increasing the precision and effectiveness of ROP diagnosis and therapy. An accurate diagnosis of ROP is essential for prompt and successful treatments to avoid vision loss in premature newborns.

However, the study is not without limitations. The dataset used in the study is limited to a single center, which may not represent other populations or regions. Additionally, the study did not compare the performance of the proposed method with that of ophthalmologists, which is an essential consideration for clinical application. Future research can address these limitations by conducting multicenter studies to validate the performance of the proposed method in other populations and comparing it with ophthalmologists' diagnostic accuracy. Moreover, future research can explore the generalizability of this approach to other retinal diseases beyond ROP. Overall, the proposed method can be a valuable tool for improving the accuracy and efficiency of ROP diagnosis, potentially leading to better treatment outcomes and improving the quality of life of premature infants.

7. References

- [1] M. Azami, Z. Jaafari, S. Rahmati, A. D. Farahani, and G. Badfar, "Prevalence and risk factors of retinopathy of prematurity in Iran: a systematic review and meta-analysis," *BMC Ophthalmology*, vol. 18, no. 1, p. 83, Apr. 2018, Doi: 10.1186/s12886-018-0732-3.
- [2] E. A. Palmer *et al.*, "Incidence and Early Course of Retinopathy of Prematurity," *Ophthalmology*, vol. 127, no. 4, Supplement, pp. S84–S96, Apr. 2020, Doi: 10.1016/j.ophtha.2020.01.034.
- [3] Aaberg T, Ben-Sira I, Charles S, et al., "An International Classification of Retinopathy of Prematurity: II," *The Classification of Retinal Detachment, Arch Ophthalmol*, vol. 105, no. 7, pp. 906–912, 1987. doi:10.1001/archophth.1987.01060070042025
- [4] H. Mahdi and N. El Abbadi, "Glaucoma Diagnosis Based on Retinal Fundus Image: A Review", *Iraqi Journal of Science*, vol. 63, no. 9, pp. 4022–4046, Sep. 2022.
- [5] M. F. Chiang *et al.*, "International Classification of Retinopathy of Prematurity, Third Edition," *Ophthalmology*, vol. 128, no. 10, pp. e51–e68, Oct. 2021, Doi: 10.1016/j.ophtha.2021.05.031.
- [6] N. Salih, M. Ksantini, N. Hussein, D. B. Halima, A. A. Razzaq, and S. A. Mahmood, "Detection of Retinopathy of Prematurity Stages Utilizing Deep Neural Networks," in *Proceedings of Seventh International Congress on Information and Communication Technology*, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., in Lecture Notes in Networks and Systems. Singapore: Springer Nature, 2023, pp. 699–706. Doi: 10.1007/978-981-19-1607-6_62.
- [7] "Progression of Type 2 to Type 1 Retinopathy of Prematurity in the Early Treatment for Retinopathy of Prematurity Study | Neonatology | JAMA Ophthalmology | JAMA Network." <https://jamanetwork.com/journals/jamaophthalmology/fullarticle/425328> (accessed Mar. 19, 2023).
- [8] Y.-P. Huang *et al.*, "Deep Learning Models for Automated Diagnosis of Retinopathy of Prematurity in Preterm Infants," *Electronics*, vol. 9, no. 9, Art. no. 9, Sep. 2020, Doi: 10.3390/electronics9091444.
- [9] A. Ellouze, O. Kahouli, M. Ksantini, H. Alsaif, A. Aloui, and B. Kahouli, "Artificial Intelligence-Based Diabetes Diagnosis with Belief Functions Theory," *Symmetry*, vol. 14, no. 10, Art. no. 10, Oct. 2022, Doi: 10.3390/sym14102197.
- [10] E. H. Wood, E. Y. Chang, K. Beck, B. R. Hadfield, A. R. Quinn, and C. A. Harper, "80 Years of vision: preventing blindness from retinopathy of prematurity," *J Perinatol*, vol. 41, no. 6, Art. no. 6, Jun. 2021, Doi: 10.1038/s41372-021-01015-8.
- [11] A. M. Freitas, R. Mörschbacher, M. R. Thorell, and E. L. Rhoden, "Incidence and risk factors for retinopathy of prematurity: a retrospective cohort study," *International Journal of Retina and Vitreous*, vol. 4, no. 1, p. 20, May 2018, Doi: 10.1186/s40942-018-0125-z.
- [12] G. M. Richter, S. L. Williams, J. Starren, J. T. Flynn, and M. F. Chiang, "Telemedicine for Retinopathy of Prematurity Diagnosis: Evaluation and Challenges," *Survey of Ophthalmology*, vol. 54, no. 6, pp. 671–685, Nov. 2009, Doi: 10.1016/j.survophthal.2009.02.020.
- [13] S. C. Chow, P. Y. Lam, W. C. Lam, and N. S. K. Fung, "The role of anti-vascular endothelial growth factor in treatment of retinopathy of prematurity—a current review," *Eye*, vol. 36, no. 8, Art. no. 8, Aug. 2022, Doi: 10.1038/s41433-021-01922-2.
- [14] J. A. F. Zupancic, G. Ying, A. de Alba Campomanes, L. A. Tomlinson, and G. Binensbaum, "Evaluation of the economic impact of modified screening criteria for retinopathy of prematurity

- from the Postnatal Growth and ROP (G-ROP) study,” *J Perinatol*, vol. 40, no. 7, Art. no. 7, Jul. 2020, Doi: 10.1038/s41372-020-0605-5.
- [15] Z. Tong, P. Xu, and T. Denœux, “An evidential classifier based on Dempster-Shafer theory and deep learning,” *Neurocomputing*, vol. 450, pp. 275–293, Aug. 2021, Doi: 10.1016/j.neucom.2021.03.066.
- [16] T. Denœux, “A neural network classifier based on Dempster-Shafer theory,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 2, pp. 131–150, Mar. 2000, Doi: 10.1109/3468.833094.
- [17] S. Lewallen, “A simple model for teaching indirect ophthalmoscopy,” *British Journal of Ophthalmology*, vol. 90, no. 10, pp. 1328–1329, Oct. 2006, Doi: 10.1136/bjo.2006.096784.
- [18] R. M. Hansen, A. Moskowitz, J. D. Akula, and A. B. Fulton, “The neural retina in retinopathy of prematurity,” *Progress in Retinal and Eye Research*, vol. 56, pp. 32–57, Jan. 2017, Doi: 10.1016/j.preteyeres.2016.09.004.
- [19] L. Brosseau et al., “Low level laser therapy (Classes I, II and III) for treating osteoarthritis,” *Cochrane Database of Systematic Reviews*, no. 3, 2004, Doi: 10.1002/14651858.CD002046.pub2.
- [20] E. D. Hansen and M. E. Hartnett, “A review of treatment for retinopathy of prematurity,” *Expert Review of Ophthalmology*, vol. 14, no. 2, pp. 73–87, Mar. 2019, Doi: 10.1080/17469899.2019.1596026.
- [21] S. Mulay, K. Ram, M. Sivaprakasam, and A. Vinekar, “Early detection of retinopathy of prematurity stage using deep learning approach,” in *Medical Imaging 2019: Computer-Aided Diagnosis*, International Society for Optics and Photonics, Mar. 2019, p. 109502Z. Doi: 10.1117/12.2512719.
- [22] Y. Tong, W. Lu, Q. Deng, C. Chen, and Y. Shen, “Automated identification of retinopathy of prematurity by image-based deep learning,” *Eye and Vis*, vol. 7, no. 1, p. 40, Dec. 2020, Doi: 10.1186/s40662-020-00206-2.
- [23] A. Ding, Q. Chen, Y. Cao, and B. Liu, “Retinopathy of Prematurity Stage Diagnosis Using Object Segmentation and Convolutional Neural Networks,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–6. Doi: 10.1109/IJCNN48605.2020.9207288.
- [24] Y.P. Huang et al., “Automated detection of early-stage ROP using a deep convolutional neural network,” *British Journal of Ophthalmology*, vol. 105, no. 8, pp. 1099–1103, Aug. 2021, Doi: 10.1136/bjophthalmol-2020-316526.
- [25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. Doi: 10.1109/CVPR.2009.5206848.
- [26] M. Byra et al., “Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images,” *Int J Comput Assist Radiol Surg*, vol. 13, no. 12, pp. 1895–1903, Dec. 2018, Doi: 10.1007/s11548-018-1843-2.
- [27] M. F. Chiang, L. Jiang, R. Gelman, Y. E. Du, and J. T. Flynn, “Interexpert agreement of plus disease diagnosis in retinopathy of prematurity,” *Arch Ophthalmol*, vol. 125, no. 7, pp. 875–880, Jul. 2007, Doi: 10.1001/archophth.125.7.875.
- [28] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” in *Proceedings of the 23rd international conference on Machine learning*, in ICML '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 233–240. Doi: 10.1145/1143844.1143874.