Smart Screening: Non-Invasive Detection of Severe Neonatal Jaundice using Computer Vision and Deep Learning

Kartikya Gupta
Intern Data Scientist
Valiance Solutions
Noida, India

Vaibhav Sharma
Data Scientist
Valiance Solutions
Noida, India

Shailendra Singh Kathait
Co-Founder & Chief Data Scientist
Valiance Solutions
Noida, India

ABSTRACT
Severe neonatal jaundice is a condition in newborns where high levels of bilirubin cause the skin and eyes to turn yellow, posing a risk of brain damage if not promptly treated. Current traditional methods are highly invasive and require intervention. Hence, this paper introduces a non-invasive approach for the preemptive detection of severe neonatal jaundice using computer vision and deep learning. The data processing pipeline includes image resizing, semantic segmentation, test split, data augmentation, and at last training-validation split. For this study, a custom CNN model was developed for binary classification alongside three transfer learning models to compare all four's performance across key metrics such as accuracy, precision, recall, F1-score, and AUC. After training, all four models were saved and then used to classify a different dataset to evaluate their performance on visually distinct images. The Vision Transformer (1.23 GiB) and EfficientNet (320 MiB) models demonstrated superior performance on testing data, achieving AUC scores of 0.87 and 0.9, respectively. However, the custom CNN model (162 MiB) and Vision Transformer achieved 0.93 and 1.0 AUC score consistently on inference data, surpassing the other models. This research contributes to creating a contactless, frugal system that can be used as a mobile application, which predicts the chances of an infant having severe jaundice.

General Terms
Neonatal Jaundice, Medical Imaging, AI in Healthcare.

Keywords
Biomedical Engineering, Bilirubin, Computer Vision, Deep Learning, Semantic Segmentation.

1. INTRODUCTION
Contrary to popular belief, neonatal jaundice [1] is a medical symptom rather than a disease itself. It is characterized by elevated bilirubin levels in newborns, resulting in the yellowing of the skin and eyes. Approximately 85% of newborns experience some degree of jaundice, which is typically mild and temporary. This condition, known as physiological jaundice, usually appears within the first few days after birth and resolves on its own as the baby's liver matures and begins to process bilirubin more effectively. However, if the infant is unable to do so, this condition can lead to significant health complications, including the risk of brain damage. Traditional diagnostic methods for severe jaundice, such as blood tests, are often invasive and may not be easily accessible, particularly in low-resource settings. These challenges highlight the need for more accessible and non-invasive diagnostic solutions.

The rapid advancements in computer vision and deep learning algorithms offer promising alternatives for disease detection. By leveraging deep learning algorithms, it is possible to develop a non-invasive frugal model that can accurately detect severe neonatal jaundice without requiring traditionally invasive detection procedures. This approach not only improves accessibility while bringing in lower cost but also enhances the efficiency of early jaundice detection. [2]

The proposed method starts with processing the images of newborns to identify jaundice. The data processing pipeline includes image resizing, semantic segmentation, data augmentation techniques, and finally training-validation split for model training. Note that after implementing the semantic segmentation model and image resizing, 25% of the data is taken as testing data, hence avoiding data augmentation of the test set while maintaining a smaller test class size of 190 images. A custom convolutional neural network (CNN) model, along with three transfer learning models —MobileNet, EfficientNet, and Vision Transformer [3]—were developed for the classification of images to predict whether an image shows a healthy infant or a jaundice. These models were evaluated on popular performance metrics like accuracy, precision, recall, F1-score, and area under the curve (AUC) to ensure comprehensive assessment and comparison.

Keeping the users in mind, inference testing of all four models becomes imperative, as their performance must be tested in conditions where the picture of an infant is taken in varying conditions like distance, lighting, camera quality, skin complexion, etc.

The study contributes to the field of medical image analysis by showcasing the efficacy of contemporary deep learning methodologies in neonatal jaundice detection. By using cutting-edge computer vision algorithms, this research aims to create a robust and frugal solution to improve clinical prognosis. Through rigorous experimentation and validation, the results show promising accuracy in jaundice detection in infants, thus enabling early intervention and improved clinical outcomes.

1.1 Severe Neonatal Jaundice
Neonatal jaundice is a common condition in newborns, characterized by the yellowing of the skin and the sclerae (the whites of the eyes) due to elevated levels of bilirubin in the blood. Bilirubin is a yellow pigment produced during the normal breakdown of red blood cells. The liver typically processes bilirubin, but in newborns, the liver is often not fully developed, leading to an accumulation of bilirubin in the bloodstream. This condition, if not managed properly, can lead

to severe complications, including kernicterus [1] [2] [3], a form of brain damage caused by very high levels of bilirubin.

1.1.1 Types

Severe neonatal jaundice can be classified primarily into two types based on the nature of bilirubin accumulation: unconjugated hyperbilirubinemia and conjugated hyperbilirubinemia.

- **Unconjugated hyperbilirubinemia (UHB):** This is the most common form of jaundice in neonates, where there is an excess of unconjugated bilirubin. This type can be physiological or pathological [5]. Physiological jaundice typically appears on the second or third day of life and resolves within a week. Pathological jaundice, however, may present within the first 24 hours and can persist longer, often indicating underlying conditions such as hemolytic diseases, genetic disorders, or internal bleeding.
- **Conjugated hyperbilirubinemia (CHB):** Less common but always pathological, this type occurs when there is an excess of conjugated bilirubin. Causes include biliary atresia, neonatal hepatitis, and metabolic disorders. Unlike UHB, CHB does not respond to phototherapy and often requires surgical intervention or specific medical treatments, depending on the underlying cause [6].
- Understanding these types helps in diagnosing the condition correctly and administering the appropriate treatment to prevent complications such as kernicterus and chronic bilirubin encephalopathy.

1.1.2 Causes

The etiology of severe neonatal jaundice is diverse, involving various genetic, environmental, and physiological factors:

- **Hemolytic Diseases:** Conditions like Rh incompatibility, ABO incompatibility, and G6PD deficiency lead to the excessive breakdown of red blood cells, overwhelming the infant's liver capacity to process bilirubin [4].
- Genetic Disorders: Inherited conditions such as Gilbert's syndrome and Crigler-Najjar syndrome [8] affect the enzymes responsible for bilirubin metabolism, leading to its accumulation [6].
- **Infections:** Neonatal sepsis and intrauterine infections can impair liver function and bilirubin metabolism, resulting in jaundice [7].
- **Prematurity:** Premature infants have underdeveloped livers, which reduces their ability to conjugate and excrete bilirubin efficiently. Additionally, they often have higher rates of red blood cell turnover, further contributing to elevated bilirubin levels [4] [5].
- **Breastfeeding:** Breastfeeding jaundice and breast milk jaundice are two forms associated with breastfeeding. The former occurs due to inadequate milk intake leading to dehydration, while the latter involves substances in breast milk that inhibit bilirubin conjugation [6].

1.1.3 Symptoms

The clinical manifestations of severe neonatal jaundice extend beyond the typical yellowish skin and sclera discoloration:

- Early Symptoms: These include jaundice appearing within the first 24 hours of life, poor feeding, lethargy, and a high-pitched cry. The early onset of jaundice is often indicative of pathological causes [4], [5].
- Progressive Symptoms: As bilirubin levels rise, symptoms can progress to include irritability, hypotonia (reduced muscle tone), and arched back with neck extension (opisthotonos) [4], [7].
- Acute Bilirubin Encephalopathy (ABE): This severe complication presents with signs such as fever, shrill cry, seizures, and even coma. ABE requires immediate medical intervention to prevent permanent neurological damage [5], [6].
- Kernicterus: Chronic and severe hyperbilirubinemia can lead to kernicterus, characterized by permanent neurological impairments including athetoid cerebral palsy, hearing loss, dental dysplasia, and cognitive deficits [4] [7].

Table 1: Bilirubin levels that may require treatment [43]

The Period Infants	Bilirubin Level
24 hours or younger infants	10 mg/dL
25 to 48 hours of infants	15 mg/dL
49 to 72 hours of infants	18 mg/dL
Older than 72 hours of, infants	20 mg/dL

1.1.4 Traditional Detection Methods

The management of severe neonatal jaundice aims to reduce bilirubin levels and prevent complications.

- Phototherapy: The primary treatment for UHB involves using blue light to convert bilirubin into a water-soluble form that can be excreted in urine. Intensive phototherapy is often required for severe cases. [4] [5].
- Exchange Transfusion: For extremely high bilirubin levels or when phototherapy is ineffective, exchange transfusion is performed to rapidly reduce bilirubin levels and remove sensitized red blood cells in cases of hemolytic disease [5].
- Intravenous Immunoglobulin (IVIG): This is used in cases of hemolytic disease of the newborn, particularly Rh incompatibility, to reduce hemolysis and bilirubin production [4].
- Treating Underlying Conditions: Identifying and managing the underlying cause is critical. This might include antibiotics for infections, hydration and increased feeding frequency for breastfeeding jaundice, or surgical intervention for biliary atresia [6].
- Supportive Care: This includes maintaining adequate hydration and nutrition, and monitoring for signs of worsening jaundice or bilirubin encephalopathy [7].

Given these challenges, there is a growing interest in developing non-invasive, automated methods for the early detection of severe neonatal jaundice. The integration of computer vision and deep learning techniques, as explored in recent studies, offers promising potential for improving the accuracy and accessibility of jaundice detection in newborns. These limitations have led to a growing interest in non-invasive detection techniques, particularly those utilizing computer vision and deep learning to analyze images of newborns' skin.

These methods can be implemented via mobile applications, leveraging the widespread availability of smartphones and digital cameras. The primary advantage of non-invasive techniques is their ability to eliminate the need for blood draws, thereby reducing discomfort and infection risk. They are also cost-effective, requiring minimal equipment and offering a scalable solution that can be integrated into existing healthcare systems. Moreover, non-invasive methods enhance accessibility, especially in rural or underdeveloped areas, by enabling early intervention and empowering parents to monitor their newborn's health without immediate access to healthcare facilities.

2. RELATED WORKS

In recent years, several studies have been conducted for the binary classification of severe neonatal jaundice from medical images. The following is a comparative analysis that looks into some of such works:

Egejuru et al. [10] developed a classification model for predicting the severity of neonatal jaundice using deep learning. They collected data on 23 variables from 23 neonatal patients at a hospital in Nigeria. The authors used a deep learning multi-layer perceptron (MLP) classifier, testing different numbers of epochs from 5 to 50. The model with 5 epochs achieved the best performance, showing the lowest mean absolute error of 0.3889. This model correctly classified 60.9% of cases overall, with perfect classification for low severity cases, 61.5% accuracy for moderate cases, and 42.9% accuracy for high severity cases. The study demonstrated that deep learning techniques can be effectively applied to classify neonatal jaundice severity, potentially improving early detection and treatment. The authors suggest that their model could be integrated into existing Health Information Systems to enhance real-time assessment of clinical information affecting the risk of kernicterus and liver disease among neonatal patients.

Nayagi et al. [11] provide a comprehensive review of computer vision techniques for detecting and classifying neonatal jaundice using color card methods. The paper examines approaches for skin detection, feature extraction, image fusion, feature selection, and classification. Key techniques include adaptive thresholding, convolutional neural networks, multi-scale image fusion, and hybrid feature selection algorithms. The authors compare classifiers like KNN, SVM, and CNN for jaundice detection, with some approaches achieving over 98% accuracy. Non-invasive bilirubin estimation methods using skin color analysis are also reviewed, with some showing a strong correlation (up to 0.83) with blood test results. The paper highlights performance metrics like accuracy, sensitivity, and specificity for various methods.

Hardalaç et al. [12] developed a mobile application for noninvasive classification of neonatal jaundice using image processing techniques. The study used images of 196 newborns, with 156 for training and 40 for testing. Multiple linear regression was applied to RGB values from 38 points on the baby's body and the color chart to estimate bilirubin levels. The system achieved 92.5% accuracy in classifying jaundice cases into two groups: <10 mg/dL and ≥10 mg/dL bilirubin. Key advantages include low processing requirements, allowing use on basic smartphones, and a simple regression model for bilirubin estimation. The authors note that this approach could aid medical professionals in rapid, noninvasive jaundice screening, especially in resource-limited settings. While showing promise, the study had a relatively small sample size and encountered some errors for bilirubin levels near the 10 mg/dL threshold.

Abdulrazzak et al. [13] conducted a study on real-time jaundice detection in neonates using machine learning models. They compared the performance of four algorithms: SVM, k-NN, RF, and XGBoost, using a dataset of 767 infant images. The XGBoost model achieved the highest accuracy at 99.63%. The authors developed a user-friendly application using MATLAB App Designer, which utilized a USB webcam for real-time image capture and analysis. The system was tested on 10 NICU patients, showing 100% agreement with the TSB test results. The study addressed the need for non-invasive jaundice detection methods, potentially reducing the use of invasive procedures. However, challenges were noted in acquiring reliable neonatal images due to factors like inconsistent lighting and limited skin tone diversity in the dataset.

The study conducted by Hashim et al. [14] focused on developing a cost-effective jaundice detection and treatment system using computer vision and color analysis. The data set included 20 infants' images from Ibn Al-Atheer Teaching Hospital, 16 internet images, and four manikin images. The methodology involved MATLAB-based image processing, including skin detection, ROI selection, and color space transformation. The system detected jaundice by analyzing skin color in the lab color space and used thresholding and morphological operations. Phototherapy was initiated with a blue LED light for detected jaundice. The models used were color-based skin detection and Otsu’s thresholding. The system accurately detected jaundice at TSB levels of 14 mg/dL and above within 1 second.

These studies collectively highlight the significant progress made in the field of neonatal jaundice classification using ML and DL models. Each paper presents unique approaches and methodologies, contributing to the overall advancement of accurate and reliable jaundice detection systems.

Table 2: Comparative Analysis of Methodologies for Neonatal Jaundice Detection

Criteria	Egejuru et al. [10]	Nayagi et al. [11]	Hardalac et al. [12]	Abdulrazzak et al. [13]	Hashim et al. [14]	Methodology
Dataset Used	Hospital data from 23 neonates	50 images of neonates with jaundice collected from a hospital	Firat University	Al-Elwiya Maternity Teaching Hospital	Ibn Al-Atheer teaching hospital	NJN Jaundice Dataset
Algorithm Implemented	Multi-Layer Perceptron-WEKA	K-NN, SVM, CNN	Multiple Linear Regression	SVM, K-NN, RF, and XGBoost	K-NN, SVR, RF,	CNN, MobileNet, EfficientNet, Vision Transformer
Methodology	10-fold cross-validation	Color card method, adaptive thresholding, multi-scale image fusion	Color calibration card, pixel similarity and white balancing methods	Regions of interest, color spaces	ROI selection, and color space transformation, Otsu's thresholding	Semantic Segmentation, Inference Testing
Performance	MAE, Accuracy, Precision, Recall	Accuracy, Sensitivity, Specificity	Accuracy	Accuracy, Precision, Recall, F1 Score, MCC	Sensitivity, specificity, and accuracy	Accuracy, Precision, Recall, F1-Score, AUC

Keys:
K-NN: K Nearest Neighbor
RF: Random Forest
SVM: Support Vector Machine
SVR: Support Vector Regression
CNN: Convolutional Neural Networks
MAE: Mean Absolute Error
AUC: Area under Curve

3. METHODOLOGY

3.1 Dataset Used

The NJN (Normal and Jaundiced Newborns) dataset [15] is a collection of images and data aimed at facilitating the diagnosis of neonatal jaundice. The dataset consists of 760 images of newborns, including 560 normal infants and 200 jaundiced infants, captured with an iPhone 11 Pro Max camera at a resolution of 1000 x 1000 pixels in JPEG format.

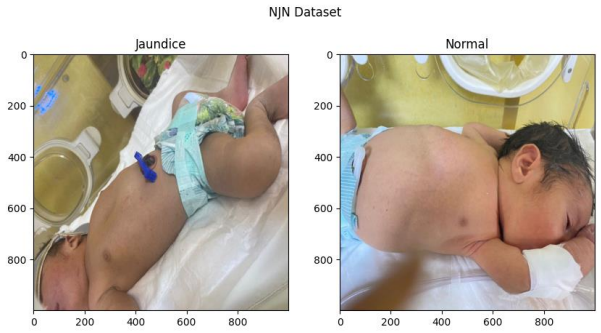

Figure 1: Sample image from NJN dataset

The images were obtained from Al-Elwiya Maternity Teaching Hospital in Baghdad, Iraq, with ethical approval and consent from legal guardians. The dataset includes newborns aged 2 to 8 days with diverse birthweights and skin tones, captured from different angles and lighting conditions. In addition to the image data, the dataset provides an Excel sheet in CSV format containing the RGB and YCrCb channel values for each image, along with the corresponding status indicating whether the newborn is normal or jaundiced. The NJN dataset aims to provide researchers and healthcare professionals with a valuable resource for training AI systems and developing algorithms for real-time, non-invasive monitoring and accurate diagnosis of neonatal jaundice. The dataset's diversity and comprehensive nature make it a significant contribution to the field of neonatal healthcare and the application of AI techniques in jaundice detection.

3.2 Data Preprocessing

The data preprocessing pipeline of the NJN jaundice dataset and started with downsizing of images from the size 1000x1000 to 224x224. This initial step helped us save on computational power and time during the innumerable experiments that undertook without compromising on the quality of research.

To maintain a rigorous testing and experimentation setup, 25% of the original dataset was partitioned as a test set, and augmentation was applied to the rest of the dataset. Following this, the models were tested, and an overfitting problem was found as the model's performance on testing data remained subpar.

To enhance testing accuracy, object detection models were explored, and various versions of the YOLO [16] models were experimented with. These efforts resulted in improved testing accuracy and overall model performance on the test data. Despite these gains, object detection images consistently underperformed on a separate inference dataset, which notably differs from the original NJN dataset. Hence, the semantic segmentation [17] approach was used. The semantic segmentation model blacked out the entire background of the infant in the original dataset so as to focus better on the baby, as any features the model picks up related to the environment would be irrelevant. Training models on images of only infants

with no background helped to train models that not only perform well under the test dataset but also when given images of infants from different settings. Augmenting these semantic segmentation images also leads to an increase in the size of the dataset, which is favorable for model training.

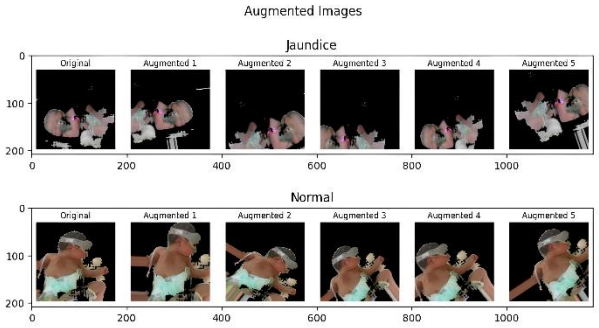

Figure 2: Data augmentation on semantic images

Algorithm 1 Semantic Segmentation [17] with LRASPP MobileNet V3 and OpenVINO

Require: Input folder *input_folder*, output folder *output_folder*

Ensure: Segmented images in *output_folder*

1: *IMAGE_WIDTH* ← 640

2: *IMAGE_HEIGHT* ← 480

3: Load segmentation model weights

4: Load OpenVINO IR model

5: **for all** images filename in *input_folder* **do**

6: **if** filename ends with .jpg, .jpeg, or .png **then**

7: *img_path* ← path to *filename* in *input_folder*

8: Read and preprocess *image* from *img_path*

9: Perform segmentation using PyTorch model

10: Convert PyTorch model to OpenVINO IR format if not done

11: Run inference on OpenVINO model

12: Extract mask for desired class (e.g., 'person')

13: Invert mask to extract background

14: Create masked image with black background

15: Normalize pixel values to range [0, 1]

16: Save masked image to *output_folder*

17: **end if**

18: **end for**

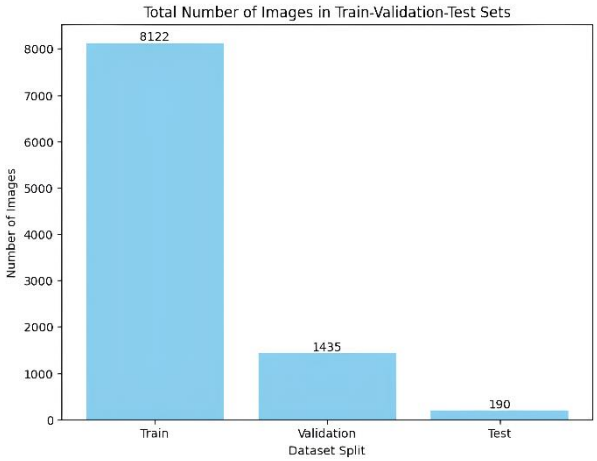

Figure 3: Data split across Train, Validation, and Test sets

3.3 Models

The following four models were implemented:

3.3.1 CNNs

A custom convolutional neural network (CNN) [18] model, which is a type of deep learning architecture commonly used for image classification tasks was developed after several iterations of trials. The model consists of various layers that process the input data through a series of convolutional layers, as shown in **Figure 4**.

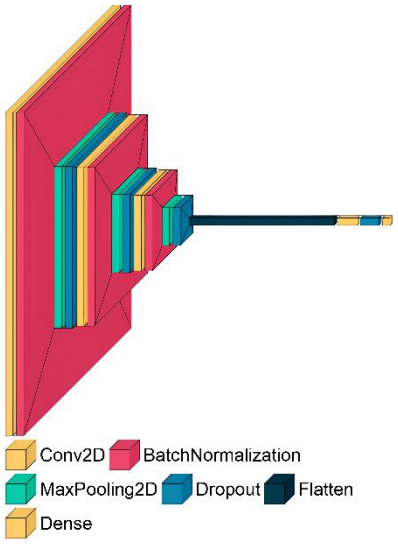

Figure 4: CNN Model Architecture

The custom CNN model depicted features multiple Conv2D layers, each followed by Batch Normalization to stabilize learning and enhance generalization. MaxPooling2D layers are employed for down-sampling, reducing spatial dimensions while retaining important features. Dropout layers are integrated to prevent overfitting, followed by a Flatten layer to convert the 2D feature maps into 1D vectors. Finally, Dense layers serve as the fully connected layers for classification. The model's superior performance on distinct inference data but suboptimal test set performance may indicate overfitting to the training-validation-test set, leading to poor generalization on unseen data of similar distribution. However, its architecture seems adept at extracting relevant features from different domains, leading to consistent exceptional performance on the divergent inference dataset.

3.3.2 MobileNet

The MobileNet V3 [19] architecture, is a convolutional neural network designed for efficient mobile and embedded vision applications. It consists of two main components: the MobileNet V3 backbone and the Segmentation Head. The MobileNet V3 backbone is responsible for extracting features from the input image at different resolutions (1/4, 1/8, and 1/16). It uses depth wise separable convolutions, which are computationally efficient, along with various optimization techniques like squeeze-and-excite blocks and hard sigmoid activations.

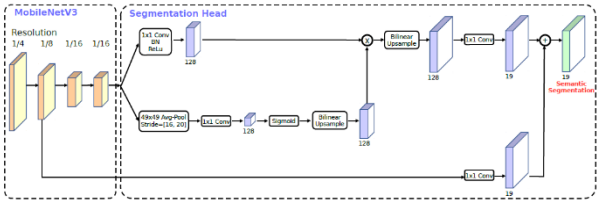

Figure 5: MobileNet V3 architecture

The Segmentation Head takes the features from the backbone and performs semantic segmentation. It consists of several convolutional layers, bilinear upsampling operations, and a sigmoid activation function to generate the final segmentation output. If fine-tuned on a specific task, MobileNet V3's architecture can potentially perform well on inference tasks due to its efficient design and the ability to capture multi-scale features. Fine-tuning the model on a target dataset allows it to adapt its learned representations to the specific task, leading to improved performance. Additionally, the segmentation head can be modified or replaced to suit the desired inference task, such as instance segmentation.

3.3.3 EfficientNet

EfficientNet [21] is a family of convolutional neural network (CNN) models designed with a focus on maximizing accuracy while maintaining model efficiency.

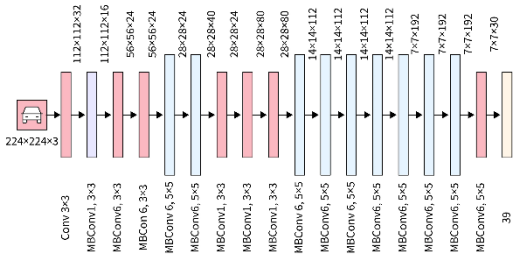

Figure 6: EfficientNet V2 architecture

The EfficientNet architecture employs mobile inverted residual blocks as the main building blocks, which are designed to be computationally efficient and effective in capturing rich representations. If fine-tuned on specific inference tasks, EfficientNet models could potentially perform well due to their ability to learn rich and transferable feature representations from the pre-training on large-scale datasets. Additionally, the efficient design of the architecture allows for relatively fast inference times, making it suitable for deployment in various applications, including semantic segmentation, and image classification.

3.3.4 Vision Transformers

Vision Transformer (ViT) [21] is a transformer-based model that applies the transformer architecture, originally designed for natural language processing tasks, to computer vision tasks such as image classification and segmentation. This architecture is a type of attention based network [22]. The ViT architecture consists of a standard transformer encoder with an additional patch embedding layer that splits the input image into fixed-size patches and then flattens and projects them into a sequence of linear embedding. These embedding are then processed by the transformer encoder, which captures long-range dependencies and generates contextualized representations for each patch.

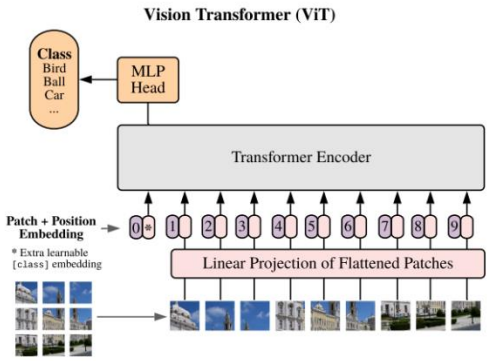

Figure 7: Vision Transformer architecture

Fine-tuning a pre-trained ViT on downstream inference tasks would conceptually lead to better performance due to the transformer's ability to effectively capture global dependencies and model long-range interactions within the image. This property allows the model to develop a comprehensive understanding of the image context, which is crucial for accurate inference tasks such as segmentation and image captioning.

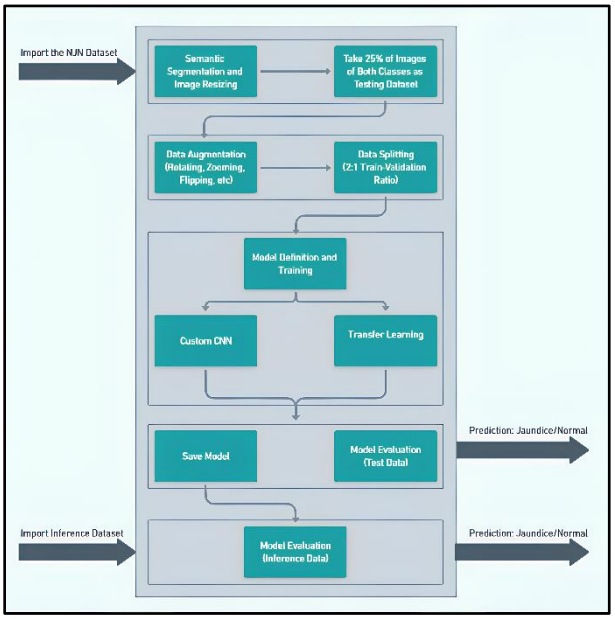

Figure 8: Flowchart of our methodology

4. RESULTS AND DISCUSSION

4.1 Testing Data

The **Figure 9** presents the confusion matrices for four different models: CNN, MobileNet, EfficientNet, and Vision Transformer. The confusion matrices provide a visual representation of the classification performance of each model by displaying the predicted class labels against the true class labels.

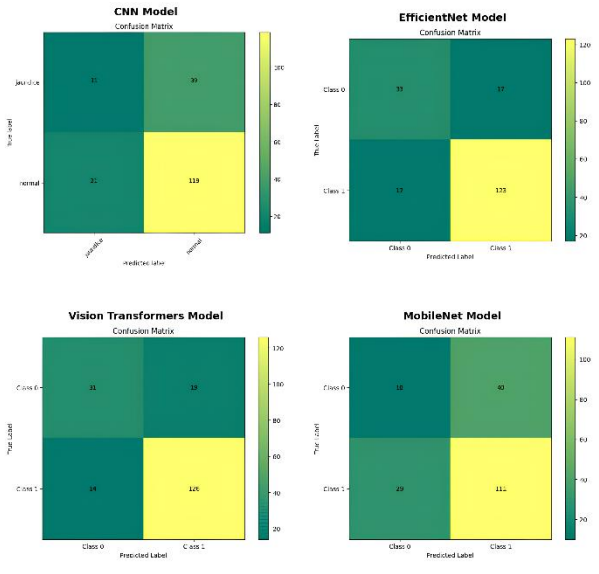

Figure 9: Confusion matrix for testing data

Overall, the confusion matrix shows a higher concentration of correct predictions (diagonal elements) for one class compared to the other, due to a class imbalance in the predictions. The transfer learning models also exhibits a similar pattern, with a higher number of correct predictions for one class over the other.

The Efficient Net and Vision Transformer models, on the other hand, appear to have a more balanced distribution of correct predictions across both classes, suggesting better overall performance in distinguishing between the two classes

Table 3: Model Evaluation on Test Data

MODEL	CLASS	A	AUC	P	R	F1-SCORE
CNN	Jaundice	79%	0.57	0.75	0.85	0.80
	Normal			0.34	0.22	0.27
MobileNet	Jaundice	64%	0.62	0.74	0.79	0.76
	Normal			0.26	0.20	0.22
EfficientNet	Jaundice	82%	0.90	0.66	0.66	0.66
	Normal			0.88	0.88	0.88
Vision Transformer	Jaundice	83%	0.87	0.87	0.90	0.88
	Normal			0.69	0.62	0.65

Keys:
- A: Accuracy
- P: Precision
- R: Recall
- AUC: Area under Curve

$$Accuracy\ (A) = \frac{TP + TN}{TP + TN + FP + FN}\ (1)$$

$$Precision\ (P) = \frac{TP}{TP + FP}\ (2)$$

$$Recall\ (R) = \frac{TP}{TP + FN}\ (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}\ (4)$$

Where,
TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

Figure 10 displays the Receiver Operating Characteristic (ROC) curves for the four models. The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The area under the curve (AUC) provides a measure of the model's overall performance, with higher values indicating better classification performance. This metric is highly valuable in the case of the used datasets here that have a class imbalance.

Accuracy simply measures the proportion of correctly predicted instances, which can be misleading here as normal class significantly outweighs the jaundice class. Hence, AUC evaluates the model's ability to distinguish between classes across different threshold settings, offering a balanced assessment by considering both true positive and false positive rates, making it more robust in imbalanced scenarios.

$$AUC = \sum_{i=1}^{n-1} \left(\frac{FPR_{i+1} - FPR_i}{2}\right) \times (TPR_{i+1} + TPR_i)\ (5)$$

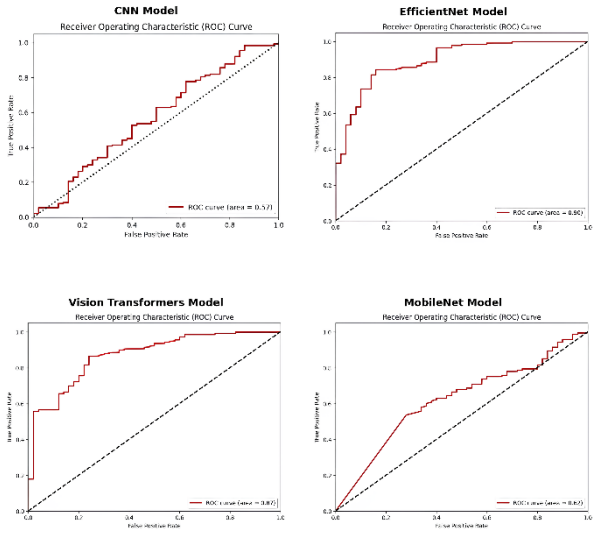

Figure 10: ROC curve for test data

Based on the ROC curves, the Vision Transformer and Mobile Net models seem to have the highest AUC values, indicating superior classification performance compared to the CNN and Efficient Net models. The CNN model appears to have the lowest AUC value, suggesting poorer classification performance relative to the other models.

4.2 Inference Data

Inference testing has always been an integral part of evaluating the performance of this methodology in real-life scenarios. During research, it is expected that the conditions under which the training data were collected will not match the conditions of the input data during real-world application. Variations in camera quality, zoom, tilt, and alignment are anticipated. To address this, images like **Figure 11** were taken from the internet to test the four models. These images were procured from a GitHub repository [23].

Figure 11: Sample image from the inference data for a) Jaundice and b) Normal class

Figure 12 provides valuable insights into the performance of different neural network architectures on an image classification task. The first image shows the inference confusion matrices for four models: CNN, MobileNet, EfficientNet, and Vision Transformer.

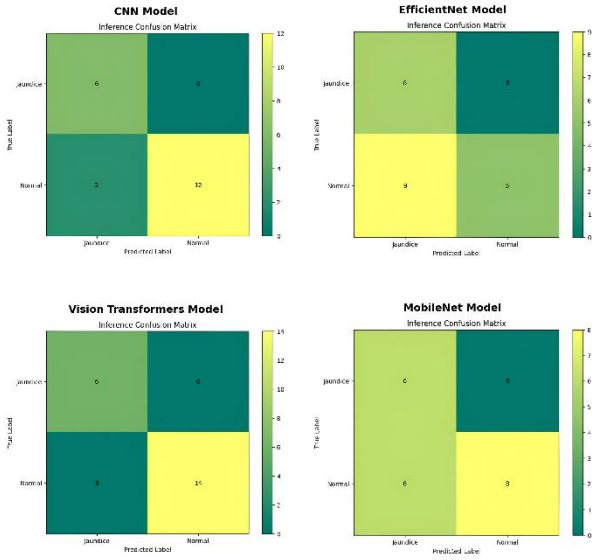

Figure 12: Confusion matrix for inference data

These matrices illustrate the models' misclassifications between the *jaundice* and *normal* classes. MobileNet and EfficientNet seem to struggle more with misclassifying normal cases as jaundice, while MobileNet and Vision Transformer have higher confusion in classifying jaundice cases as normal.

The **Figure 13** presents the Receiver Operating Characteristic (ROC) curves for the same models. ROC curves are useful for evaluating the trade-off between true positive rate and false positive rate at different classification thresholds.

Interestingly, the ROC curves for all four models are quite similar, with the CNN having significantly better performance, as indicated by their higher Area under the Curve (AUC) values. On the other hand, the EfficientNet that performs well on the test set fails to classify inference images.

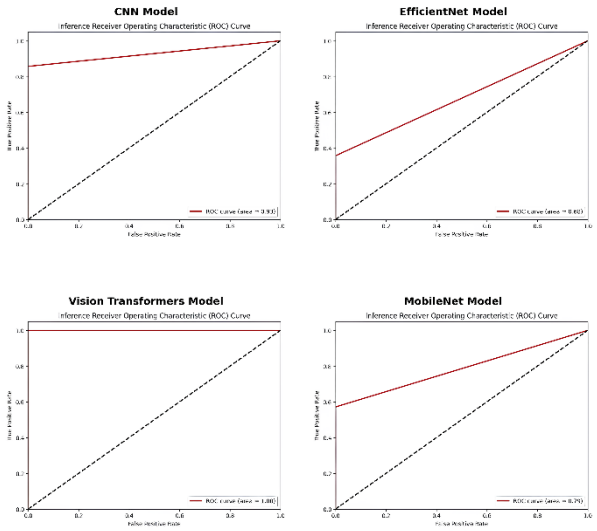

Figure 13: ROC curve of Inference Data

Table 4: Model Evaluation on Inference Data

MODEL	CLASS	A	AUC	P	R	F1-SCORE
CNN	Jaundice	90%	0.93	1.00	0.86	0.92
	Normal			0.75	1.00	0.86
MobileNet	Jaundice	70%	0.62	1.00	0.57	0.73
	Normal			0.50	1.00	0.67
EfficientNet	Jaundice	55%	0.68	1.00	0.36	0.53
	Normal			0.40	1.00	0.57
Vision Transformer	Jaundice	100%	1.00	1.00	1.00	1.00
	Normal			0.69	0.62	0.65

Keys:
- A: Accuracy
- P: Precision
- R: Recall
- AUC: Area under Curve

5. CONCLUSION

In this study, various deep learning models were explored to enhance the early detection of neonatal jaundice using medical imaging. It was observed that while the attention-based ViT model gives stellar performance on both testing and inference data, it is a clearly larger model (1.22 GiB), which cannot qualify as a frugal and lightweight solution. The CNN model shows immense potential due to its persistent performance on the inference dataset, which is a unique observation of its own.

This research opens up the possible implementation of TinyML techniques like pruning, knowledge distillation, and even the Lottery Ticket Hypothesis. A lighter version of these models could potentially be integrated into clinical workflows to aid in the timely diagnosis and treatment of neonatal jaundice. Valiance Analytics is looking into building an end-to-end solution that might not be limited to use by clinical experts but also by a common man. Work on a larger dataset for both model training and inference is also desired, subject to availability.

6. ACKNOWLEDGEMT

The NJN Jaundice Dataset, provided by Abdulrazzak et. Al [13] was invaluable for this research on newborn jaundice detection.

7. REFERENCES

[1] Slusher, T.M., Zamora, T.G., Appiah, D., Stanke, J.U., Strand, M.A., Lee, B.W., Richardson, S.B., Keating, E.M., Siddappa, A.M. and Olusanya, B.O., 2017. Burden of severe neonatal jaundice: a systematic review and meta-analysis. BMJ paediatrics open, 1(1).

[2] Althnian, A., Almanea, N. and Aloboud, N., 2021. Neonatal Jaundice Diagnosis Using a Smartphone Camera Based on Eye, Skin, and Fused Features with Transfer Learning. Sensors, 21, p.7038.

[3] Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y. and Yang, Z., 2022. A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence, 45(1), pp.87-110.

[4] Diala, U.M., Usman, F., Appiah, D., Hassan, L., Ogundele, T., Abdullahi, F., Satrom, K.M., Bakker, C.J., Lee, B.W. and Slusher, T.M., 2023. Global Prevalence of Severe Neonatal Jaundice among Hospital Admissions: A Systematic Review and Meta-Analysis. Journal of Clinical Medicine, 12(11), p.3738.

[5] Kinshella, M.L.W., Salimu, S., Chiwaya, B., Chikoti, F., Chirambo, L., Mwaungulu, E., Banda, M., Hiwa, T., Vidler, M., Molyneux, E.M. and Dube, Q., 2022. Challenges and recommendations to improve implementation of phototherapy among neonates in Malawian hospitals. BMC pediatrics, 22(1), p.367.

[6] Ansong-Assoku, B., Shah, S.D., Adnan, M. and Ankola, P.A., Neonatal Jaundice.

[7] Trasancos, C. and Horey, D., 2024. Experiences with neonatal jaundice management in hospitals and the community: interviews with Australian health professionals. BMJ open, 14(2), p.e075896.

[8] Tcaciuc, E., Podurean, M. and Tcaciuc, A., Management of Crigler-Najjar syndrome. world, 5, p.6.

[9] Labpedia (n.d.) 'Liver function Tests:- Part 4 – Neonatal Jaundice Classification and Diagnosis', Labpedia, accessed 14 June 2024. Available at: https://labpedia.net/liver-function-tests-part-4-neonatal-jaundice-classificationand-and-diagnosis/

[10] Egejuru, N.C., Asinobi, A.O., Adewunmi, O., Aderounmu, T., Adegoke, S.A. and Idowu, P.A., 2019. A classification model for severity of neonatal Jaundice using deep learning. American Journal of Pediatrics, 5(3), pp.159-169.

[11] Nayagi, S.B. and Angel, T.S., 2022. Detection and Classification of Neonatal Jaundice Using Color Card Techniques--A Study. International Journal of Online \& Biomedical Engineering, 18(15).

[12] Hardalac, F., Aydin, M., Kutbay, U.Ğ.U.R.H.A.N., Ayturan, K., Akyel, A., Çağlar, A., Hai, B. and Mert, F., 2021. Classification of neonatal jaundice in mobile application with noninvasive imageprocessing methods. Turkish Journal of Electrical Engineering and Computer Sciences, 29(4), pp.2116-2126.

[13] Abdulrazzak, A.Y., Mohammed, S.L., Al-Naji, A. and Chahl, J., 2024. Real-Time Jaundice Detection in Neonates Based on Machine Learning Models. BioMedInformatics, 4(1), pp.623-637.

[14] Hashim, W., Al-Naji, A., Al-Rayahi, I.A., Alkhaled, M. and Chahl, J., 2021. Neonatal jaundice detection using a computer vision system. Designs, 5(4), p.63.

[15] Abdulrazzak, A.Y., Mohammed, S.L. and Al-Naji, A., 2023. NJN: A Dataset for the Normal and Jaundiced Newborns. BioMedInformatics, 3(3), pp.543-552.

[16] Reis, D., Kupec, J., Hong, J. and Daoudi, A., 2023. Real-time flying object detection with YOLOv8. arXiv preprint arXiv:2305.09972.

[17] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N. and Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), pp.3523-3542.

[18] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.

[19] Howard, A., Sandler, M., Chu, G., Chen, L.C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V. and Le, Q.V., 2019. Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324).

[20] Tan, M. and Le, Q., 2021, July. Efficientnetv2: Smaller models and faster training. In International conference on machine learning (pp. 10096-10106). PMLR.

[21] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.

[23] Sardana, A. (n.d.). 1-Neonatal Jaunice Detection/Dataset/Face. GitHub. Available at: https://github.com/AshishSardana/jaundice-detection/tree/master/1-Neonatal%20Jaunice%20Detection/Dataset/Face.