New Born Baby Jaundice Detection Using Time Slot Scheduling With GA Based Feature Selection

1Arulmozhi, A and 2Dr. Ezhilarasi, M
1Avinashilingam University, Department of Printing Technology, Assistant Professor, Box.3030, Coimbatore, Tamilnadu, India.
2Kumaraguru College of Technology (KCT), Department of Electronics and Instrumentation (E & I), Professor & Head,Box.3030, Coimbatore, Tamilnadu, India.

ABSTRACT

Background: New Born Baby Jaundice Detection Using Time Slot Scheduling With GA Based Feature Selection. Objective: Jaundice is the yellowish discoloration of the skin and conjunctive of the eyes. In this analysis, a Genetic Algorithm (GA) is used to enhance or optimize the overall behavior by evolving the population. Genetic Algorithms (GAs) are a virtually new criterion for a search, based on the precept on instinctive selection. The main intent of this phase is to reduce the time consumption by using time slot scheduling and management. An ensemble of classifiers is a predetermined classifier whose individual decisions are combined in some way to classify new examples. In this analysis, an ensemble of fitness evaluation would produce an ensemble of fitness values for each individual. In this phase, the works are shared by resources to optimize feature selection time. The experimental results reveal that the proposed method can act as a supplement to support earlier detection and more effective treatment due to improved jaundice detection.

Results: In this analysis, the accuracy level was increased by 100%. In this paper, the time consumption is reduced by using time slot scheduling and management.

Conclusion: Jaundice is one of the most familiar conditions demanding medical attention in newborn babies. Jaundice refers to yellow coloration of the skin and the sclera and is induced by a constructed level of bilirubin in the circulation, a position recognized as hyperbilirubinaemia. In this research, preprocessing is done by using the hybrid median filtering technique and GLCM is used for texture feature extraction and color feature extraction. In this analysis, an ensemble of fitness evaluations would produce an ensemble of fitness values for each individual. In this paper, the work allocation problem has been tackled with Genetic Algorithm (GA). The worker allocation is a combinational optimization problem and the contribution of the worker allocation is an automated tool based on GA with a time slot scheduling algorithm that can be used to assign work to workers. Finally, the Kernel Support Vector Machine (SVM) is used to classify the normal babies and jaundice detected babies.

INTRODUCTION

Jaundice is a yellow discoloration of the skin and whites of the eyes that is often seen in newborn babies. The discoloration is caused by a yellow substance called bilirubin. Infants with high blood levels of bilirubin, called hyperbilirubinaemia, develop the yellow color when bilirubin accumulates in the skin. Jaundice is not a disease, but is a symptom of an elevated bilirubin levels are not handled in a timely manner. Jaundice is caused by raised levels of bilirubin the body, a condition known as hyperbilirubinaemia. Jaundice is caused by raised levels of bilirubin in the body, a condition known as hyperbilirubinaemia. The discoloration is caused by a yellow coloring of the skin and conjunctive of the eyes.

Jaundice can also make babies sleepy, which can lead to poor feeding. Jaundice is common in newborn babies and it occurs as a result of the liver being underdeveloped and not fully functional. Bilirubin comprises an open chain of four pyrrole-like rings (tetrapyrrole). In heme, these four rings are linked into a larger ring, termed a porphyrin ring. Bilirubin is mainly produced from the breakdown of red blood cells. Red cell breakdown produces unconjugated (or indirect) bilirubin, which circulates mostly in bound form with albumin.
although some are freely and hence able to enter the brain. In newborns, jaundice is detected by stabilizing the skin with digital pressure so that it concedes underlying skin and subcutaneous tissue. Mild jaundice is natural in newborn baby, which normally disappears within a few days as the enzymes are formed in the body. Texture is an innate property of all surfaces that describes visual patterns, each having homogeneity. Feature extraction is carried out by using colors, using textures or by using shapes. Image texture can be seen as an image area containing repeated patterns of pixel intensities arranged in some structural way.

Gray Level Co-occurrence Matrix (GLCM) is a popular representation of the texture in images. It contains a count of the number of times a given feature occurs in a special spatial relation to another given feature. GLCM, one of the most known texture analysis methods, estimate image properties related to second order statistics. Color is the most extensively used “feature” owing to its intuitiveness compared with other features and most importantly, it is easy to extract from the image. Color, local shape, texture and spatial information in a variety of forms are the most extensively used features in such systems. Genetic Algorithms (GAs), a form of inductive learning strategy, are adaptive search techniques which have demonstrated substantial improvement over a variety of random and local search methods. The main issues in applying GAs to any problem are selecting an appropriate representation and an adequate evaluation function. In the feature selection problem the main interest is in representing the space of all possible subsets of the given feature set. Choosing an appropriate function is an essential step for successful application of GAs to any problem domain. In order to use GAs as the search procedure, it is essential to define a fitness function which properly assesses the decision rules generated by the machine learning algorithm. Ensemble methods are very popular in machine learning. The obstacle of classification in machine learning consists of using labeled examples to induce a model that classifies objects into a set of known classes.

The machine learning community has only attacked the problem of “optimal” feature selection indirectly in that the traditional biases for simple classification rules leads to efficient induction procedures for producing individual rules containing only a few features to be evaluated. The main objective of this analysis is to reduce the time consumption by using time slot scheduling and management. The feature selection work is distributed to N-workers and the time of work execution is calculated by using distributed function evaluation. Support vector machines (SVMs) use a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high dimensional feature space.

The remaining part of the paper is organized as follows: Section II involves the works related to probable solutions for newborn infants jaundice detection. Section III involves the description of the proposed method – Newborn baby jaundice detection through the optimized feature set. Section IV involves the performance analysis of the proposed work. The paper is concluded in Section V.

Related work:
Qi, et al suggested a feature selection method for multivariate performance measures (Qi and Tsang, 2013). Feature selection with precise multivariate performance measures was the key to success of many applications such as text classification and image retrieval. The existing feature selection methods were designed for classification error. Prasad and Singh recommended a study on blood bilirubin levels in a tertiary care center. Jaundice occurs in most newborn infants (Prasad, 2014). Bilirubin is primarily produced from the breakdown of red blood cells. Ali, et al designed an algorithm for diagnosis of the three kinds of constitutional jaundice (Ali, 2010). Constitutional jaundice is a rare disease which has little influence on either the well-being or longevity of the patient, it commonly has a good prognosis. In this paper, the proposed algorithm was designed to diagnose the three types of constitution jaundice namely, Gilbert, Rotor and Dubin-Johnson syndrome. (De Carvalho, 2011) described the clinical approach to term and near term newborn infants with severe hyperbilirubinemia. Arun Babu, et al described the association between peak serum bilirubin and neurodevelopmental outcomes in term babies with hyperbilirubinaemia (Babu, 2012). Neonatal hyperbilirubinaemia is an imperative cause of preventable brain damage and early death among infants. Yousef, et al suggested a study using one class machine learning for miRNA target discovery (Yousef, 2010). Machine learning enables one to generate automatic rules based on observation of the appropriate examples of the learning machine. Support Vector Machines (SVMs) are a learning machine developed as a two class approach.

Mansor, et al implied a PCA based feature extraction and k-NN algorithm for early jaundice detection (Mansor, 2011). The intent of this research was the development of a stand-alone automated system that could be used as a supplement in the NICU to provide 24-h/day noninvasive. In this paper, the recommended jaundice detection algorithm localizes the face of the given input image using Haar Classifier method. Abdelgader, et al aimed to measure the hematological parameters in newborn children to detect the frequency of anemia among newborns (Abdelgader, 2013). All 80 newborns were included in this study. All babies were term, 78 of them were healthy and while 2 were diseased. Own and Abraham contemplated a weighted rough set framework for early intervention and prevention of neurological dysfunction and kernicterus that are catastrophic sequels of neonatal jaundice (Own and Abraham, 2012). Rough set theory was an adequately disparate intelligent technique that has been applied to the medical domain and was used for the discovery of data dependencies. In
this paper. They applied class equal sample weighting to build a weighted information table. (Onyearugha, 2011) determined the occurrence, aetiological and other associated factors of neonatal jaundice in medical center. The overwhelming majority of the inborn babies, 90.4% in this study were exclusively breastfed and were significantly more than 12.7% of the out born babies fed likewise.

Sambath and Debjit (Kumar, 2011) suggested a review of clinical features, differential diagnosis and remedies. They described that, jaundice is not an illness, but a medical condition in which too much bilirubin a compound produced by the breakdown of hemoglobin from red blood cells was circulating in the blood. Jaundice was due to a buildup in the blood of bilirubin, a yellow pigment which comes from the breakdown of old red blood cells. It is normal for red blood cells to break down, but the bilirubin formed does not commonly cause jaundice because the liver metabolizes it and gets rid of it into the gut. (Kekre, et al 2010) implied a different method for image retrieval based on texture feature extraction using Vector Quantization (VQ). In this paper, they used Linde-Buzo-Gray (LBG) and Kekre’s Proportionate Error (KPE) algorithms for texture feature extraction. The identification of specific textures in an image was achieved primarily by modeling, texture as a 2D gray level variation. This 2D array is called as Gray Level Co-occurrence Matrix (GLCM). (Mansor, et al,2012) suggested a color detection method for detection of jaundice in newborn infants. A simple color detection method was employed to study the behavior of the infants. In newborns jaundice was detected by blanching the skin with digital pressure so that it reveals underlying skin and subcutaneous tissue.

Chen, et al., designed an Self Organization Map (SOM) algorithm and its application to color feature extraction. Quality of features extracted or data reduced by the conventional SOM algorithm heavily depends on the distribution of the training data. (Nassef, et al., 2013) evaluated G6PD activity and antioxidants status in jaundiced Egyptian neonates. Jaundice with glucose-6-phosphate dehydrogenase (G6PD) deficiency is one of the most common conditions needing medical attention in newborn babies. This study included 100 infants, 40 with neonatal jaundice and 60 jaundice free cases. This analysis showed an incidence of hyperbilirubinemia, G6PD deficiency and oxidative stress biomarkers changes among jaundiced newborns in Egypt. Laddi, et al investigated a non-invasive and instant method of jaundice detection using machine vision technique (Laddi, A., 2013). The proposed technique was developed based upon non-invasive detection of jaundice by using simple image acquisition and processing tools. (Bisoi, et al., 2012) suggested a study to find out the occurrence of G6PD deficiency and its association with relevant clinical, gestational, socio-demographic factors. In this study, it was observed that only 5% babies had parents of consanguineous marriage. Out of those 5%, 1% were G6PD deficient.

Zhao, et al designed a new approach based on genetic algorithm with feature chromosomes (Zhao, 2011). Support Vector Machines (SVMs) were an emerging data classification technique with many diverse applications. In this paper, the asymptotic behaviors of SVM were fused with GA. (Jagadeesh, et al., 2012) suggested a genetic algorithm approach for singular value decomposition. In this paper, the proposed method is based on quantization step size optimization using the genetic algorithm to enhance the quality of watermarked image and robustness of the watermark. Balin addressed parallel machine scheduling problems with fuzzy processing times (Balin, 2011). In this paper, a robust GA approach embedded in a simulation model was proposed to diminish the maximum completion time.

Proposed method:

The intent of this analysis is to reduce the time consumption by using time slot scheduling and management. The time of the feature selection process is optimized by sharing of resources to N-workers. The whole work process is distributed to N-workers and the distributed function evaluation are used to calculate the time of the worker processes. The development of the jaundice detection system involves the following tasks:

- Preprocessing
- Feature Extraction
- Feature Selection
- Time slot scheduling and Management
- Work Distribution to N-Workers
- Distributed Function Evaluation
- SVM Classification

A) Preprocessing:

In this analysis, a novel filter structure is used, namely. Hybrid Median Filter, which constitutes a natural extension of the nonlinear rational type hybrid filters. A technique for processing an image, comprises a foreground and a background to yield a greatly compressed and specific portrayal of the image. In hybrid median filtering, the neighboring pixels are ranked according to brightness (intensity) and the median value becomes the new value for the central pixels. In the median filtering operation, the pixel values in the neighborhood window are ranked according to the intensity and the middle value (median) becomes the output
value for the pixel under evaluation. The hybrid median filter is more expensive to compute than a smoothing filter.

Given a set of random variables $M = (M_1, M_2, \ldots, M_N)$, the order statistics $M_{(1)} \leq M_{(2)} \leq \ldots \leq M_{(N)}$ are random variables, defined by sorting the values of $M_i$ in an increasing order. The median value is then given as,

$$\text{Median}(M) = \begin{cases} M_{(\lfloor f \rfloor + 1)}, & \text{for } X = 2A + 1 \\ \frac{1}{2} (M_{(i)} + M_{(i+1)}), & \text{for } X = 2A \end{cases}$$

Where, $f = 2A + 1$ is the median rank. The median is considered to be a robust estimator of the location parameter of a distribution and has found numerous applications in smoothing. For a grayscale input image with intensity values $x_{i,j}$, the hybrid median filter is defined as,

$$Y_{i,j} = \text{median}_{(a,b) \in G}(x_{i+r,j+s})$$

Where, $G$ is a window over which the filter is applied. The symmetric square window of size $A \times A$ with $A = 2C + 1$, i.e., the median rank $f$ equals $f = (A^2 + 1)/2$. This is probably also the most widely used form of this filter.

**B) Feature Extraction:**

Texture is one of the significant characteristics used in identifying objects or regions of interest in an image. GLCM is one of the analytical methods of scrutinizing texture that considers the spatial relationship of pixels in the GLCM, also known as the gray level spatial dependence matrix. The GLCM functions are used to describe the texture of an image by calculating how often pairs of pixels with distinct values in a particular spatial relationship that occurs in an image. This created GLCM is then used for extracting statistical measures. GLCM is a second order statistical feature which incorporates information about pixels having related gray level values in an image. Texture is an imperative aspect of an image that has been extensively used in medical image classification, remote sensing and visual classification.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables commonly requires a large amount of memory and computation power or a classification algorithm which ever fits the training sample and generalizes poorly to new samples. The GLCM method is a way of extracting second order statistical texture features. The features of GLCM are,

1. Contrast
2. Correlation
3. Cluster prominence
4. Cluster shade
5. Dissimilarity
6. Energy
7. Entropy
8. Homogeneity
9. Local homogeneity
10. Maximum Probability
11. Sum of Squares
12. Auto correlation

For extraction of color, texture features, RGB and HSV color spaces are used. HSV (Hue, Saturation and Value) color space is a nonlinear transformation of the RGB color space. Feature vector is computed for every Hue (H), Saturation (S) and Value (V) channel:

$$FV = [\text{mean (H)}, \text{mean (S)}, \text{mean (V)}, \text{variance (H)}, \text{variance (S)}, \text{variance (V)}, \text{skewness (H)}, \text{skewness (S)}, \text{skewness (V)}]$$

Where, $FV$ – Feature Vector. GLCM describes the frequency of one gray tone appearing in a specified spatial linear relationship with another gray tone, within the area under investigation.
Input Dataset → Preprocessing → Feature Extraction → Share the Feature Selection Work → Distributed Function Evaluation

Scheduling and Management

Worker 1
Worker 2
Worker 3
Worker n

Whether all iterations have completed?

No → Yes

Save the Results

Fig. 1: Newborn baby jaundice detection by using a feature selection framework with time slot scheduling.

C) Feature Selection:

Feature selection is one of the key topics in machine learning, it can remove the irrelevant, even noisy features and hence improve the quality of the data set and the performance of learning systems. Applying Genetic Algorithms (GAs) to the feature selection problem is straightforward: the chromosomes of the individuals contain one bit for each feature and the value of the bit determines whether the feature will be used in the classification. Using the wrapper approach, the individuals are evaluated by training the classifiers. Ensemble methods are learning algorithms that construct a set of classifiers and then classify new data points by taking a vote of their predictions.

The main discovery is that ensembles are often much more accurate than the individual classifiers that make them up. In this paper, a recent and generic machine learning algorithm based on decision tree ensembles that has been shown to perform remarkably well on a variety of tasks. Feature selection gives the chance to attain the same or better performance by using fewer features. So, two terms are adequate in fitness function, accuracy and the number of selected features. With these parameters the fitness function should be maximized during optimization operation.

\[
\text{Fitness function} = \text{Error} + \mu \text{Ones}
\]

(4)

Where, Error corresponds to the classification error for a specific subset of features and Ones corresponds to the number of features selected. \( \mu \) is a coefficient that expresses the weight of second parameter of fitness function (Ones) versus the first one (Error). A necessary and sufficient condition for an ensemble of classifiers to be more accurate than any of its individual members is if the classifiers are accurate.
D) Time slot Scheduling and Management:

Scheduling can be described as the obstacle of assigning work to a given set of workers in such a way that some constraints are satisfied. Time slots represents the sequence of the processing of the detailed works at each stage, where the work assignment to each slot and its associated processing times are variables to be determined. At each stage in the process, there can be single or multiple units, and when multi units are involved, time slots are defined for each unit. The scheduling system can use various kinds of parameters in order to arrive at a scheduling policy to optimally complete an execution. Time slot scheduling is being increasingly regarded as an essential requirement in high performance systems. In this paper, an important subset of schedulers namely slot schedulers that discretize time into quanta called slots. Slot schedulers are generally used for scheduling works in a large number of applications. In this analysis, a more flexible approach called time slot scheduling is proposed. A slot scheduler treats time as a discrete quantity. It splits time into discrete quanta called slots. By using a time slot scheduling algorithm the time consumption is reduced for the overall feature selection process.

E) Time Slot Scheduling Algorithm:

The algorithm of time slot scheduling with GA based feature selection is shown below,

Step 1: Select images from input data set.
Step 2: Hybrid Median Filter is used to preprocess the set of images.
Step 3: After that, the GLCM based feature extraction is done.
Step 4: GA with time slot scheduling algorithm is applied after selecting the features from the set of images.
Step 5: In feature selection,

Define a feature matrix \( F = [F_{ij}] \) as follows:

For all \( (i, j) \in E \), set \( F_{ij} = 0 \) and for all \( (i, j) \in E \) set \( F_{ij} = F_i + \delta_{ij} \).

Where, \( F_i \) represents feature separation, \( \delta_{ij} \) is a randomly chosen number from the interval \((0, \text{total number of workers})\) and can be selected by one communication between i, j.

Step 6: The algorithm variables are updated training features that exchange between neighboring nodes.

Let \( F^{0}_{i \rightarrow j} \in \mathbb{R} \) denote updated training feature from i to j in iteration k.

Step 7: Initialize, \( k = 0 \) and set the updated training feature as follows,

\[
F^{0}_{i \rightarrow j} = F_{ij}, \quad F^{0}_{j \rightarrow i} = F_{ij}
\]  

Step 8: After selecting the features, SVM classifier is used to classify the normal and abnormal images.

F) Work distribution to N-workers:

The distribution of work to workers is based on capabilities of workers. In this phase, the feature selection work is distributed to N-workers. For the successful application of workflow technology, it is crucial that there be a good fit between work practice and the models used by the workflow management system. Work flow processes are operational processes where there is a highly dynamic trade-off between security and performance. Genetic Algorithm (GA) is considered a promising method so in this paper work distribution problem is solved by using GA. It provides basic genetic mechanisms that can be easily used to apply evolutionary principles to problem solutions. In specific, it can be seen that the GA always vastly enhanced the fitness of the solution from the beginning to the end. Distribution function evaluation is used to combine the results of workers. After feature selection, the resultant images will be classified by using the SVM classifier.

G) SVM Classification:

The Support Vector Machine is a theoretically superior machine learning methodology with great results in pattern recognition. The function can be a classification function or the function can be a general regression function. The kernel based method is based on mapping data from the initial input feature space to a kernel feature space of higher dimensionality and then determining a non-linear problem in that space. These techniques allow us to interpret learning algorithms geometrically in the kernel space, thus incorporating statistics and geometry in an efficient way.

Performance Analysis:

A) Preprocessing:

Fig 2. (A) depicts the input image. Baby image 16 is given as the input image. The newborn danger signs are defined as the signs which occur in the newborn within 30 days of life, such as pathological jaundice, poor feeding and excessive weight loss. Fig 2. (B) shows the skin detection. The preeminent demand in skin detection is to constitute the recognition robust to the large variations in appearance of skin that may occur, like shape, intensity, color, etc. Skin color and texture are important cues that people use consciously or unconsciously to infer variety of culture related aspects about each other. Skin detection is the process of finding skin colored pixels and regions in an input image.
Fig 2. (C) represents the region of interest. A region of interest is a portion of an image that performs the filter operation on the image. Fig 2. (D) illustrates the preprocessed image. In this work, the hybrid median filtering technique is used for preprocessing. The aim of preprocessing is an advancement of the image data that conceals unwanted distortions or enlarges some image features essential for further processing.

![Image](A) ![Image](B)

![Image](C) ![Image](D)

**Fig. 2:** Skin detection and Preprocessing.

**B) Genetic Algorithm:**

Fig 3 shows the genetic algorithm with fitness evaluation. In this graph, the x-axis represents the generation and the y-axis represents the fitness value. The average fitness of the whole population is the fitness of each genotype multiplied by its frequency, this is known as mean fitness.

![Graph](Genetic Algorithm)

**Fig. 3:** Genetic Algorithm.

**C) Confusion Matrix:**

Fig 4. Shows the confusion matrix. The confusion matrix function allows comparison of a classified image. A confusion matrix contains information about actual and predicted classifications done by a classification system. The performance of such systems is generally classified using the data in the matrix.
Fig. 4: Confusion Matrix.

D) **Analysis:**

The comparison between the normal babies and jaundice affected babies are shown in Fig 5. In this analysis, the classification rate and the level of accuracy are increased by using the ensemble tree. Kernel SVM classifier is used to classify the normal babies and the jaundice detected babies.

Fig. 5: Overall Analysis.

E) **Time Comparison:**

Fig 6 shows the time comparison between the existing and proposed methods. In this graph, the x-axis represents the existing and proposed methods and the y-axis represents the time in seconds. In this analysis, the time consumption is reduced by using time slot scheduling and management.

Fig. 6: Time comparison between existing and proposed methods.
**F) Performance Measures:**

The performance measure for time slot scheduling is shown in Table 1. In this analysis, the accuracy level was increased by 100%. In this paper, the time consumption is reduced by using time slot scheduling and management.

**Table 1: Performance Measure for time slot scheduling.**

<table>
<thead>
<tr>
<th>Performance</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Rate</td>
<td>1</td>
</tr>
<tr>
<td>Error Rate</td>
<td>0</td>
</tr>
<tr>
<td>Last Correct Rate</td>
<td>1</td>
</tr>
<tr>
<td>Last Error Rate</td>
<td>0</td>
</tr>
<tr>
<td>Inconclusive Rate</td>
<td>0</td>
</tr>
<tr>
<td>Classified Rate</td>
<td>1</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>1</td>
</tr>
<tr>
<td>Specificity</td>
<td>1</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>1</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>1</td>
</tr>
<tr>
<td>Positive Likelihood</td>
<td>NaN</td>
</tr>
<tr>
<td>Negative Likelihood</td>
<td>0</td>
</tr>
<tr>
<td>Prevalence</td>
<td>0.7500</td>
</tr>
</tbody>
</table>

**Conclusion and Future Work:**

Jaundice is one of the most familiar conditions demanding medical attention in newborn babies. Jaundice refers to yellow coloration of the skin and the sclerae and is induced by a constructed level of bilirubin in the circulation, a position recognized as hyperbilirubinaemia. In this research, preprocessing is done by using the hybrid median filtering technique and GLCM is used for texture feature extraction and color feature extraction. In this analysis, an ensemble of fitness evaluations would produce an ensemble of fitness values for each individual. In this paper, the work allocation problem has been tackled with Genetic Algorithm (GA). The worker allocation is a combinational optimization problem and the contribution of the worker allocation is an automated tool based on GA with a time slot scheduling algorithm that can be used to assign work to workers. Finally, the Kernel Support Vector Machine (SVM) is used to classify the normal babies and jaundice detected babies.

**REFERENCES**


