iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

INNOVATIVE AI-BASED TECHNIQUES FOR NON-INVASIVE ANEMIA SCREENING AND DIAGNOSIS

P. Vasudha,
S. Banu Vamsi,
T. Reshwanth,
B. Tech Students, Dept. of Computer Science and Engineering,
R.V.R & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Dr. S J R K Padminivalli V
Associate Professor, Dept. of Computer Science and Engineering,
R.V.R & J.C College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

ABSTRACT
Because many cases of anemia are still found too late, particularly in areas with limited access to laboratory testing and medical resources, this article developed. Early diagnosis is further hampered by costly testing procedures and intrusive technologies. To identify anemia, a non-invasive, precise, and efficient technique is needed. The conjunctival picture of the eye is examined in this study as a non-invasive way to identify anemia. In an effort to effectively classify anemic and healthy patients, a number of model techniques were attempted. The Support Vector Machine (SVM) algorithm-integrated MobileNetV2 method was determined to be the most effective plan. With this combination, the accuracy of 93%, sensitivity of 91%, and specificity of 94%. These findings show that the model can successfully identify healthy patients while accurately identifying anemic patients. This method offers a non-invasive means of detecting anemia early on, making it promising for use in clinical settings. The SVM+MobileNetV2 technique relies on images of the eye's conjunctiva and can potentially improve healthcare by identifying people who may have had earlier anemia. This technique stands out as a solid option for the efficient and precise diagnosis of anemia when accuracy, sensitivity, and specificity are balanced.

Keywords:
Anemia detection, eye conjunctival images, MobileNetV2, non-invasive, SVM.

INTRODUCTION

A medical disease known as anemia occurs when the level of red blood cells in the bloodstream is below normal [1]. Red blood cells, also called erythrocytes, transfer oxygen from the lungs to various tissues and organs and carry carbon dioxide, a waste product, away from those tissues to be expelled through the lungs [2]. This helps the body function as a whole. According to epidemiological data, anemia, which affects over a third of the global population, is the most prevalent blood illness [3]. The detection process is also a complex challenge. WHO highlights that it is still difficult to accurately and thoroughly detect anemia everywhere globally, particularly in regions with developed economies. Anemia has a significant global impact, with considerable implications for social and economic growth [4]. Each government must examine the expenditures of combating the disease, which includes preventive, traditional analytical approaches, administration, and lodging services [5]. Estimating the financial and social consequences of anemic patients is very problematic. Some patients are frequently required to undergo recurrent laboratory tests, which might be inconvenient due to blood samples and raise transportation and/or support costs. Anemia can be detected and diagnosed using a variety of procedures available to medical professionals [6]. These approaches aim to analyze the composition and quality of blood components, notably red blood cells, and their associated characteristics. comprehensive overview of the different types of blood cells, including red blood cells, white blood cells, and platelets [6]. Another technique, known as a peripheral smear, involves the microscopic examination of a blood sample's cellular components,

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

especially red blood cells [7]. This can reveal distinctive patterns or abnormalities in the cells' size, shape, and appearance, which can provide valuable insights into the presence of anemia and its underlying causes [8]. The reticulocyte count is another important diagnostic tool [9]. Reticulocytes are young, immature red blood cells released into the bloodstream from the bone marrow [10]. Serum iron indices, including parameters like serum iron, total iron-binding capacity (TIBC), and transferrin saturation, offer information about iron levels and utilization in the body [11]. Iron deficiency is a common cause of anemia, and assessing these indices can help identify whether anemia is related to insufficient iron availability [12]. However, some techniques require the unpleasant procedure of blood collection, which might make some people anxious or distressed. This emphasizes how crucial non-invasive substitutes are [13]. By looking at the body's exterior signals, doctors can learn important information about possible anemia situations. Due to low oxygen levels, anemia patients may exhibit pallor in their hands, nails, tongue, and eyes. Particularly, the pallor of the eyes might offer important signs of anemia. Thus, it is crucial to create non-invasive, affordable, and conveniently accessible anemic diagnosis techniques [13]. Notably, machine learning (ML) approaches have been successful in a variety of domains, most notably medicine, where they have been applied to the analysis of anemia. [14], [15], [16], [17], where K-Nearest Neighbor (KNN), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM) are examples of popular machine learning techniques. Although KNN is simple to use, it is sensitive to outliers and susceptible to changes in the data. Although Naïve Bayes is notable for its quick and simple performance, its flaw is its naive feature independence assumption, which isn't always true. Although RF prevents overfitting and works well with big datasets, it might be hampered by difficult-to-understand model complexity and longer computation times. SVM performs well in high-feature spaces, manages model complexity, and effectively addresses overfitting. All things considered, SVM is an effective and adaptable technique for categorizing a variety of datasets, including imbalances [18], [19], [20], and [21]. SVM can also be used as a classifier layer for deep learning (DL) models. The recognition performance of DL models can even be enhanced by employing SVM [22], [23]. One DL technique that is frequently used for image recognition or classification is the convolutional neural network (CNN). Performance will be impacted by the CNN model, and its design requires expertise. Pre-trained models are CNN models that may be used as a foundation for a variety of image recognition tasks after being taught different patterns and features from sizable datasets. Benefits include overcoming data restrictions, utilizing information from vast datasets, and effectively transferring learning. There are numerous well-known pre-trained models available, including VGG, DenseNet, Inception, ResNet, NasNet, Xception, MobileNet, and others. [23], [24], [25], [26], [27], and [28]. A MobileNet variation called MobileNetV2 is a tiny, highly computationally efficient model created especially for mobile applications. In addition to its ability to generate high accuracy, its benefits include a customizable architecture, depth wise separable convolution for enhanced efficiency, and scaling multipliers for flexibility in modifying model size according to application needs. This study significantly addresses some of these outstanding issues by introducing a revolutionary machine learning-based technique that is non-invasive and economical to enable the automatic detection of anemia. Promising outcomes were obtained from a training and testing procedure on the provided hardware and software solutions using conjunctival images from the Eyes defy-anemia simulation dataset collection. To address the problem of imbalance in class data, the methodology employed in this work combines two approaches: Tomek Links [29] and the Synthetic Minority Over-sampling Technique (SMOTE) [30]. The goal of this project is to create an automated, non-invasive, and reasonably priced anemia diagnosis device based on machine learning. The number of samples in the minority class was increased by creating synthetic samples from anemic patients' conjunctival pictures using the SMOTE technology. In the meantime, nearby conjunctival image pairs between anemic and healthy patients are removed using the Tomek Links technique. By merging these two processes, the conjunctival image data set became more evenly distributed between the two groups. The focus has switched to applying the three techniques recommended in this work—SVM, MobileNetV2, and the MobileNetV2-SVM combination—in the anemia diagnosis process once the processes to address the class imbalance were successfully completed. SVM is a method for classifying data that can identify intricate patterns. SVM was able to differentiate between photos of healthy conjunctiva and those that suggested anemia by using balanced data.

RELATED WORKS

Many academics have been concentrating on creating non-invasive techniques to detect anemia in recent years. Among these is the application of machine learning algorithms, which are becoming more and more common

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

for identifying and diagnosing clinical conditions. For instance, Asare et al.'s study employed machine learning to identify iron deficient anemia. To find out which algorithm is better at detecting anemia in children, researchers examined the color of the nails, the palpability of the palms, and the conjunctiva of the eyes using Nave Bayes, CNN, SVM, k-NN, and decision trees. This work was divided into three phases: data preparation, dataset collection, and model creation for the diagnosis of anemia. The results show that CNN has the highest accuracy (99.12%) and SVM has the lowest (95.4%). These findings show that a non-invasive technique can effectively identify anemia [33]. In a different study, Mannino et al. employed a novel, non-invasive technique to diagnose anemia. They used artificial intelligence (AI) techniques and smartphones to conduct a remote operation that allows for rapid haemoglobin level screening. Using this method, the conjunctiva is automatically extracted as a region of interest (ROI) from an eye shot. In order to train machine learning algorithms to determine whether a person is anemic, the next step is to analyze the ROI and extract important factors. After testing on over 200 individuals, this model has an accuracy of 85%, a precision of 86%, and a recall of 81%. These findings imply that a non-invasive method that makes use of AI and smartphone technologies may be helpful for accurately and quickly identifying anemia. [34] Dimauro's research has developed a revolutionary machine learning-based intelligent system that can automatically diagnose anemia using image analysis of the conjunctiva of patients from Italy and India. In this work, anemia was diagnosed by the machine learning algorithm using two different areas of the mucous membrane of the eye. After being extensively trained using palpebral conjunctival pictures, the RUSBoost algorithm has shown excellent performance in classifying patients into those who have anemia and those who do not [13]. The study's foundation is the publicly accessible Eyes-Defy-Anemia dataset, which includes 218 eye photos that were taken using a particular method to reduce the effects of ambient light. Particularly in the scleral segmentation test, intriguing findings were obtained with good precision (88.53), recall (82.53), and F1 scores (84.10). In the anemia identification task, the F2 score was 83.8% when employing the vessel color feature alone and 86.4% when employing the color feature of the entire sclera [30]. Furthermore, it seemed that the haemoglobin value and the color feature correlated nicely. In a separate study, Dimauro and Simone proposed a technique they named graph partitioning segmentation. Using this technique, semantic segmentation will be done in areas that have a particular diagnostic significance.

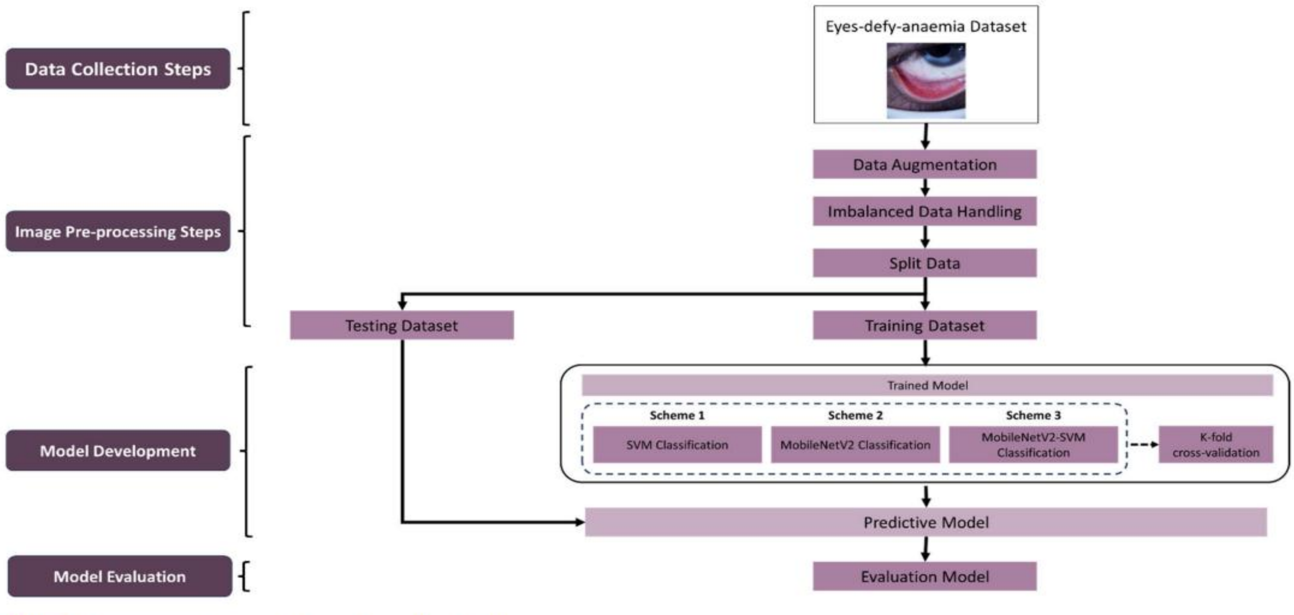

FIGURE 1. Pre-processed images obtained from the original image.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

When compared to earlier approaches, the combined method of SMOTE and Tomek Links may yield better results in terms of accuracy and generalizability. Furthermore, the usage of SVM, MobileNetV2, and the MobileNetV2-SVM combination in anemia identification is the main emphasis of this study. It suggests addressing this problem by combining conventional approaches with educational strategies. Because the MobileNetV2 algorithm, which has benefits in visual object recognition, and the data set used may make this study plan very different from other techniques. By leveraging state-of-the-art technology and combining many algorithms and approaches to overcome class imbalances, this research has the potential to considerably contribute to the development of more accurate and efficient anemia detection tools.

METHODOLOGY

The "Eyes-defy-anemia" dataset is used as the data source for the first phase of data collection, which is one of the various sections of the study framework (Figure 1). Segmented pictures of Italian and Indian patients' eyes are included in this dataset. To get the data ready for a machine learning application, data pretreatment procedures including conjunctival image augmentation were then carried out. Additionally, this study paradigm reconciles the disparities in sample sizes between the groups of anemic and normal patients using the Tomek Links and SMOTE (Synthetic Minority Over-sampling Technique) techniques [36], [37]. The data is separated into two subsets—training and testing—after being balanced. The split uses the label stratification approach with a 70:30 ratio to guarantee that the classes are fairly represented in both subsets [38], [39]. Following preprocessing, the data is ready for the modeling phase, which includes feature extraction and data standardization necessary for machine learning. Three different modeling techniques are used in this study framework: MobileNetV2, Support Vector Machine (SVM), and the MobileNetV2-SVM combination. Every technique has a unique mechanism for gathering and categorizing data from pictures of the conjunctiva of the eye. After the modeling procedure, a model evaluation is carried out to see how well each method detects anemia. A range of evaluation criteria were used to gauge the effectiveness of these strategies.

A. DATASET AND PRE-PROCESSING DATA

The dataset from this work, Eyes-defy-anemia [13], significantly advances efforts to diagnose anemia by looking at the eyes' pale conjunctiva. Images of Indian and Italian patients' eyes taken using smartphones that have been adjusted to fit devices are included in this dataset. The majority of the eye area is suitably exposed by this shooting technique. From these eye images, the palpebral or whole (palpebral and forniceal) conjunctival regions were subsequently recognized and manually distinguished. 211 picture samples make up this dataset after certain samples were eliminated based on predetermined standards (Figure 2). It consists of identically sized segmented eye photos (1067 × 800 pixels).

FIGURE 2. Example image of the data set used in this study.

The two patient subgroups in this dataset are Indian and Italian patients. Images of 123 Italian patients' eyes were obtained from their blood samples. Nevertheless, we must exclude several samples from our analysis [13], [30]. Because their forniceal conjunctiva, a specific portion of the eye, was not visible in pictures, patients 1, 35,

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

54, 58, 75, and 109 were excluded from the study. Patient 93 was also removed due to a decrease in Hb concentration. Thus, the Italian patient group was analyzed using 116 ocular pictures. Conversely, 95 blood samples and eye pictures from Indian patients were collected at Karapakkam, Chennai, India, and are all identified as Indian data sets [13]. Two categories were created from 211 eye image samples, 42 of which came from anemic patients and the remaining 169 from healthy patients [13].
The collection include not only pictures and segmentation data but also other important data including the patient's age and gender, as well as laboratory-measured Hb values. These factors all work together to produce more thorough assessments and high-caliber, perceptive research. This dataset can be used to design and test machine learning algorithms for detecting anemia based on conjunctival pallor and identify connections between clinical factors, including Hb value, age, and gender.

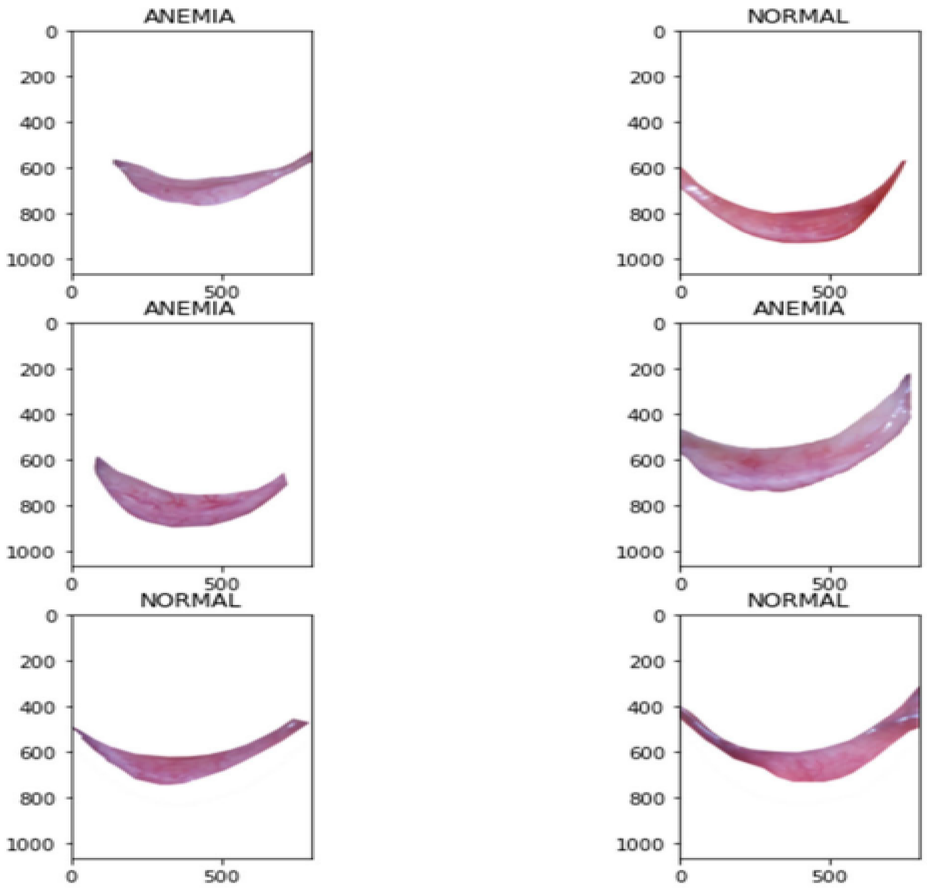

FIGURE 3. Pre-processed images were obtained from the original image.

In this case, the conjunctival photos undergo a number of modifications to provide a more varied and rich dataset. One of the enhancement techniques is the resize procedure [42]. The conjunctival picture initially has three color components (R, G, and B) and 800x1067 pixel dimensions. Nevertheless, the image is downsized to 224 × 224 pixels with three color components as part of the augmentation process. Because of this scaling, the image is smaller and easier to analyze using machine learning algorithms. Next, normalize the pixel values in every image [43]. The image's pixel values range from 0 to 1 when each pixel value is divided by 255.0. Moreover, image duplication is used to provide data augmentation. Three times the initial data is multiplied by the initial data. Rotation and flip are two types of augmentation used in this process [46]. The pictures will be flipped and rotated at an angle between -30 and 30 degrees. Following the data augmentation stage, the label encoder approach is applied [47]. The dataset's class labels will all be transformed

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

into numerical values to aid in the training of machine learning models. The study's labels fall into two categories: "normal" and "anemia." To make model processing easier, the number 0 will stand for the label "anemia," and the number 1 for the label "normal." Machine learning algorithms can comprehend the category class label when it is converted into a numeric form by a label encoder [48]. This is known as a label encoder, and it entails transforming categorical class labels into a machine-readable numeric format. During training, the model is ensured to accurately comprehend and identify the target class by encoding labels [49], [50]. Before adding data to the machine learning model, this step is essential because it fills the gap between textual class representations and the numerical representations needed by machine learning techniques. Figure 3 shows the result of this preprocessing step, showing how picture data is converted into a matrix format suitable for machine learning analysis.

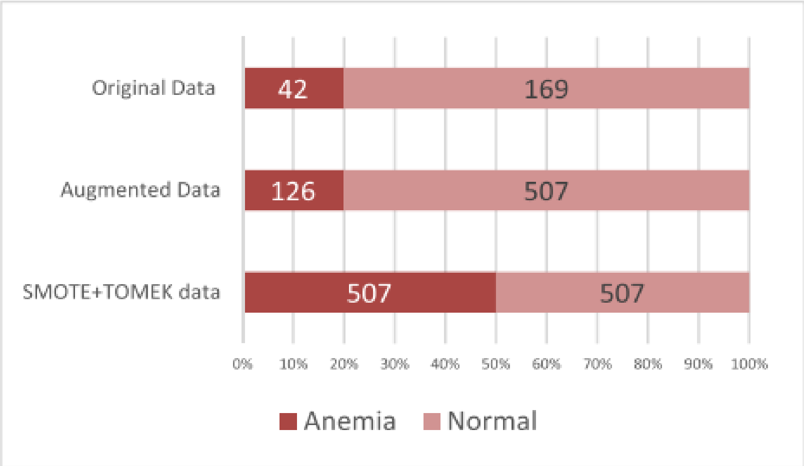

FIGURE 4. Composition of the dataset before and after handling the imbalanced data.

In order for machine learning models to learn from a variety of image orientations and views, this augmentation aims to increase the dataset's diversity. [51]. A common technique in image data processing to increase the quantity and quality of data used in machine learning processes is augmentation. In order to enable machine learning algorithms to learn from a variety of scenarios that can arise in the real world, augmentation enhances the previously utilized data set with various visual alterations [52]. Through this augmentation procedure, 42 anemic patients and 169 normal people become 126 anemic patients and 507 normal people. This data will undergo additional processing, specifically the imbalance data.

B. HANDLING IMBALANCED DATASETS

Dealing with data imbalance comes next. Upon identifying that the expanded dataset still contained an imbalance—126 anemia patients and 507 normal patients—the next step was to address this issue. To address class imbalance, this work employed Tomek Links and SMOTE (Synthetic Minority Over-sampling Technique) [36], [53]. SMOTE is an oversampling technique that expands the number of samples in the minority class by using synthetic data derived from preexisting samples. This method creates synthetic data by removing the sample from the nearest neighbor and multiplying it by a random number after selecting a random sample from the minority class and determining its nearest neighbor [37]. Conversely, the goal of the under-sampling technique Tomek Links is to reduce the number of samples in the majority class. This method identifies sample pairs that are the closest neighbors to one another while having differing class names. Furthermore, the Tomek Links pair sample is removed from the majority class [54].
The SMOTE + Tomek strategy, which involves utilizing SMOTE to oversample the minority class and Tomek Links to eliminate samples from the majority class that are Tomek Links pairs, is used in this work in combination with both of these approaches. It combines SMOTE and Tomek Links in an effort to balance the dataset's classes. Figure 4 displays the results of the class imbalance handling procedure. It is evident from these results that the dataset's class composition has stabilized, with 507 for the anemic class and 507 for the normal

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

class. This treatment approach produced a balanced distribution between the two classes, which would have enhanced the machine learning model's ability to detect anemia with greater accuracy. With a balanced composition, the model is more likely to understand and forecast the two groups accurately. To ensure that the finished model can correctly detect anemia in patients, this is an essential step to do.

C. SPLIT DATASET PROCESS

Following data pre-processing and data imbalance management, the stratify label approach is used to divide the data into training and testing subsets at a 70:30 ratio [38], [39]. A critical first step is preparing data for machine learning models. Training data (train) and test data (test), two distinct subsets of previously processed data, are separated by the Train-Test Split. The model will be trained using 70% of the data, and its efficacy will be tested using the remaining 30%, according to the 70:30 ratio that is being used. Training and testing subsets of the data are separated using the stratified label technique. This suggests that the distribution of classes in the original dataset will be preserved in both selections.

D. PROPOSED ALGORITHM MODEL FOR CONJUNCTIVAL CLASSIFICATION

SVM, MobileNetV2, and the combination of MobileNetV2-SVM are the three methods. In Figure 5, the flow of each scheme is displayed. In this context, "method" refers to the approach or algorithm used to identify and classify features in eye conjunctival image data. The feature extraction approach is essential because it extracts and transforms an image's key characteristics into a numerical format that machine learning algorithms can comprehend. SVM, MobileNetV2, and MobileNetV2-SVM combinations are some of the methods employed in this instance. Each of these techniques will be explained in depth in the sections that follow, along with how they are applied to modeling for anemia diagnosis. This will provide you a comprehensive understanding of each scheme's specific procedures and how they help identify anemic people.

1) CLASSIFICATION USING SVM ALGORITHM (SCHEME 1)

Anemia in ocular conjunctival image data was identified using the SVM technique with a linear kernel in Scheme 1 of this study [56]. The linear kernel SVM algorithm was used because it can divide two data classes linearly, which is appropriate when the pattern of separation between the classes can be accurately represented in the appropriate feature space by a straight line [57]. SVM is used in the classification method, and linear kernels to produce a linear "line" or "hyperplane" that divides the two data classes as efficiently as possible [58]. This line is positioned so that it is the furthest from points of various classes (also known as support vectors) [59]. The fundamental idea of SVM is kept. But the SVM model uses a linear kernel to find the optimal linear dividing line between different classes [60], [61]. PCA (Principal Component Analysis) techniques are also used for dimension reduction in order to help visualize how SVM with a linear kernel separates the two classes in a condensed feature space [62], [63]. Two essential features that sufficiently explain data fluctuation are obtained by condensing feature dimensions using PCA. By visualizing the dimension reduction results in a two-dimensional plane, we can see how SVM with a linear kernel projects data into a smaller space.

2) CLASSIFICATION USING MobileNetV2 ALGORITHM (SCHEME 2)

In Scheme 2 of this work, anemia detection in eye conjunctival image data was classified using the MobileNetV2 algorithm. A convolutional neural network (CNN) architecture called MobileNetV2 was developed mainly for image identification applications [64], [65]. MobileNetV2's benefit is that it can effectively extract features using lighter convolution procedures, making it suitable for use in low-power devices like mobile phones [66]. A feature extraction step using a fine-tuned CNN architecture is the first step in the MobileNetV2 classification process [67]. In order to distinguish between anemia indications and normal photos, MobileNetV2 will automatically evaluate the crucial components of the conjunctival image data [68], [69]. In order for the model to find pertinent patterns connected to various classes, the network's parameters will be changed during the training phase [68]. For a typical item recognition task, the MobileNetV2 algorithm in Scheme 2 is pre-configured and trained on vast volumes of visual data, frequently from the ImageNet dataset. At this stage, though, the model will be modified and retrained using pre-processed and unbalanced eye image datasets.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

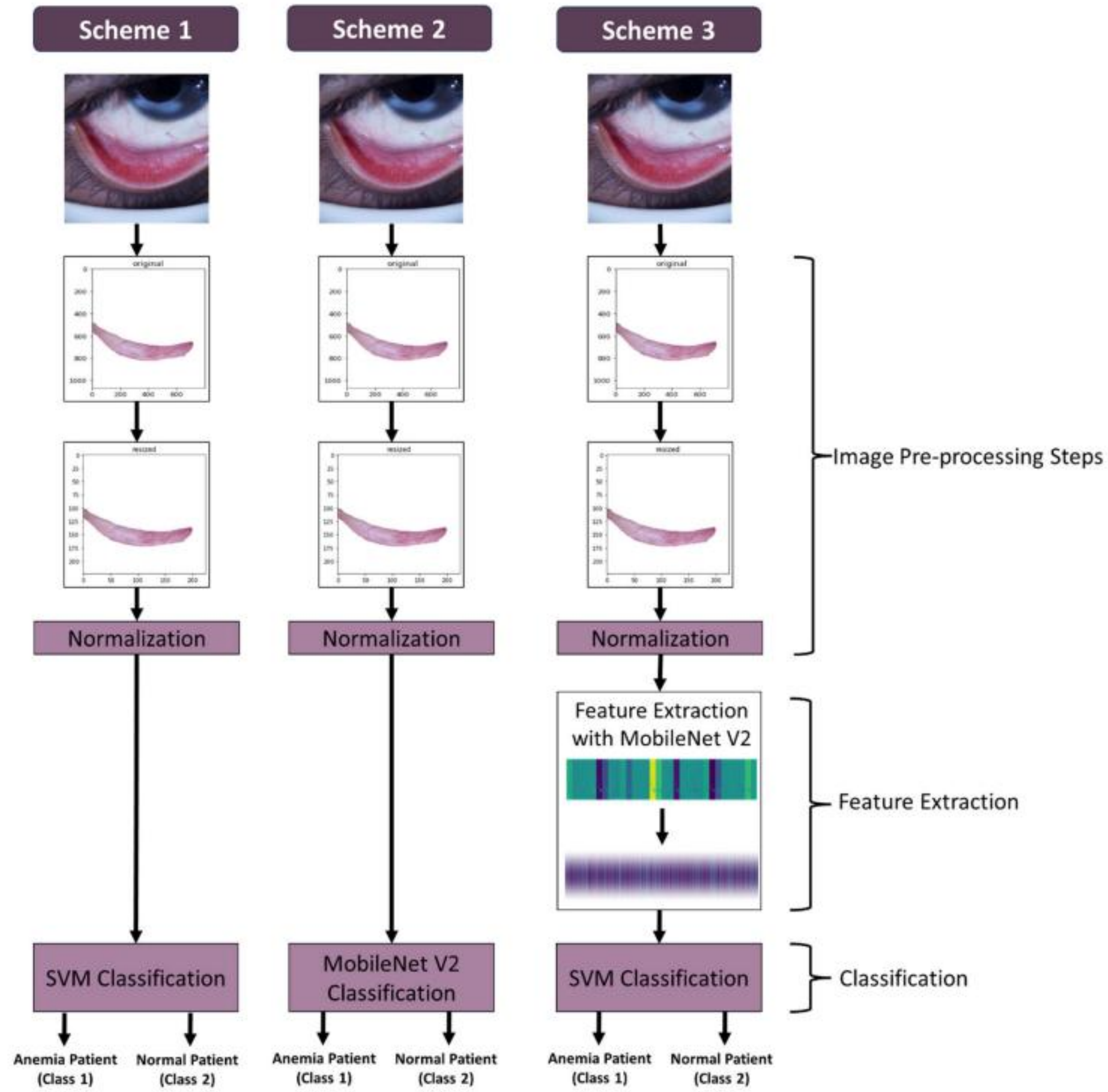

FIGURE 5. Algorithm flowchart for each scheme.

Certain visual characteristics associated with anemia can now be identified by the MobileNetV2 model on ocular conjunctival pictures. Scheme 2 illustrates the potential application of the MobileNetV2 algorithm as a classification tool for identifying anemia in datasets of eye images. Utilizing MobileNetV2's proficiency in pattern recognition and feature extraction, this method is customized for the particular data in this study.

3) CLASSIFICATION USING SVM+MobileNetV2 ALGORITHM (SCHEME 3)

In order to classify and diagnose anemia in ocular conjunctival image data, Scheme 3 used the SVM (Support Vector Machine) and MobileNetV2 algorithms. Using the benefits of each technique, this combination aims to address classification issues and optimize anemia detection results. Using the MobileNetV2 method, the SVM+MobileNetV2 combination classification process starts the feature extraction step [70]. The MobileNetV2 model will look for visual characteristics associated with anemia symptoms by analyzing conjunctival images. The SVM algorithm is then fed the results of the feature extraction process. The features that MobileNetV2 collected will be used by SVM to build a classification model. SVM's primary objective is to create an ideal

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

"line" or "hyperplane" between the two classes in order to differentiate between normal and anemic conjunctival pictures [56]. This pairing makes it possible to leverage SVM's ability to create strong class boundary separations and MobileNetV2's deep feature extraction capabilities from images [71]. A more reliable classification model that can predict anemia detection with greater accuracy is produced by this combination. This work uses Scheme 3 in an attempt to combine the advantages of the SVM and MobileNetV2 algorithms into a single model that can identify anemia signs on conjunctival images of the eye. Combining the performance of anemia diagnosis could be significantly improved by using a variety of approaches.

The K-fold or K-fold cross-validation approach is used to validate all schemes during the modeling process, as seen in Figure 1. Evaluating the suggested model's performance in each scheme is the goal of this validation [72], [73]. Using varying k values for cross-validation, the K validation-fold technique separates data into groups or "folds" [73]. The testing data for each iteration will be one of the folds, and the training data will be the remaining k-1 folds. Each fold is used as both test and training data, therefore this process is done k times. To reduce variances and generate consistent and trustworthy results, K-fold validation needs to be carried out correctly. K-fold validation is primarily used to provide a more consistent assessment of model performance along with trustworthy data on training errors. In this experiment, k = 5 was used in the K-fold validation technique [74], [75].

E. EVALUATION MODEL

At the conclusion of the modeling procedure, the model evaluation stage looks at how well the model detects anemia in the dataset of ocular conjunctival images [76]. In order to better understand how the model behaves when classifying data, this evaluation makes use of a number of crucial criteria. The metrics used include specificity, sensitivity, and accuracy [77].

1) Accuracy(PRECISION)

The model's total capacity for accurate prediction is evaluated by accuracy. Accuracy in this study will show how well the model categorizes eye conjunctival pictures as either anemic or generally normal. This is essential for assessing the overall performance of the model [77].

2) TRUE POSITIVE RATE (SENSITIVITY)

Sensitivity evaluates the model's capacity to detect actual positive results. Sensitivity will indicate how well the model can identify pictures of the eye's conjunctiva displaying anemic symptoms in the context of this study [19], [78]. To guarantee that the model can detect patients who might have anemia, sensitivity is needed.

3)SPECIFICITY (TRUE NEGATIVE RATE)

This metric evaluates the model's capacity to detect actual negative cases [79]. In this case, specificity shows how effectively the model can identify pictures of the eye's conjunctiva that are actually normal. To prevent the model from wrongly classifying healthy patients, specificity is necessary.

With the use of these indicators, we may obtain a more comprehensive picture of model performance. While sensitivity and specificity help us understand a model's ability to independently distinguish between positive and negative classifications, accuracy gives us a basic idea of how well a model can predict. The results of this study, which are based on the modeling scheme previously discussed, will help determine how well the proposed model recognizes and separates conjunctival images of the eye showing signs of anemia from normal ones.

RESULT AND DISCUSSIONS

A. EXPERIMENT RESULT

The goal of this research is to develop a conjunctival image-based anemia detection system using the three suggested model systems. This sub-chapter will address the experimental findings of the three model schemes that were previously suggested. The outcomes of this experiment shed light on how well each model performed in classifying pictures of the eye's conjunctiva as normal or anemic. The three model schemes' experimental results will provide light on how well each model detects anemia symptoms in pictures of the conjunctiva of the eye. In order to detect anemia in pre-processed image data, this will help determine the effectiveness and applicability of each model technique.

1) CONJUNCTIVAL IMAGE CLASSIFICATION USING THE SVM ALGORITHM (SCHEME 1)

In the initial classification, this work uses a linear kernel using the SVM technique. An SVM function that works with data meant for linear classification is called a linear kernel.

Every kind of kernel in SVM has unique characteristics, and additional parameters called C or cost parameters need to be changed in the case of a linear kernel. By optimizing parameter C, the linear kernel on SVM looks for

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

the optimal line of separation between data. This parameter establishes the trade-off between decreasing test data misclassification (to avoid underfitting) and permitting more training data misclassification (to prevent overfitting). The linear kernel process optimizes parameter C to achieve the highest classification accuracy. This means utilizing a trial-and-error method or testing a large number of models with different C values. The classification performance measure, which is frequently assessed using accuracy measures, can be used to assess the results of parameter C optimization using a linear kernel. In order to identify anemia from photographs of the conjunctiva of the eye, this work uses a linear kernel in SVM to find the parameter C that offers the best classification accuracy.

TABLE 1. The results of adjusting parameter C's value in Scheme 1 on the testing dataset.

Fold	PARAMETER C	Accuracy	Sensitivity	Specificity
Fold 1	1	0,8732	0,7465	1,000
	10^{-1}	0,8732	0,7465	1,000
	10^{-2}	0,8662	0,7465	0,986
Fold 2	1*	0,8873	0,8451	0,930
	10^{-1}	0,8873	0,8451	0,930
	10^{-2}	0,8873	0,8451	0,930
Fold 3	1	0,8873	0,8028	0,972
	10^{-1}	0,8873	0,8028	0,972
	10^{-2}	0,8732	0,7746	0,972
Fold 4	1	0,8662	0,8028	0,930
	10^{-1}	0,8662	0,8028	0,930
	10^{-2}	0,8662	0,8028	0,930
Fold 5	1	0,8794	0,7714	0,986
	10^{-1}	0,8794	0,7714	0,986
	10^{-2}	0,8794	0,7714	0,986

*Optimal parameter C value in Scheme 1.

Conjunctival images can be accurately classified as normal or anemic using the SVM model with a linear kernel, depending on the optimal choice of parameter C.The initial classification's experimental results utilizing the SVM method with a linear kernel are shown in Table 1.Now, in order to enhance classification performance on training and testing data, researchers look at different values of parameter C. Parameter C was tested with three different values in this experiment: 10−2, 10−1, and 1. Model performance was assessed based on a number of factors, such as specificity, sensitivity, and accuracy. The trade-off between allowing training data to be misclassified and minimizing test data misclassification is controlled by parameter C. When tested on test data, the model with the best C parameter performs best on the second cross-validation, with an average accuracy of 88.73%. There was a 93% specificity and an 84.5% sensitivity. The ability of the SVM model with a linear kernel to identify anemia in conjunctival images is demonstrated by this, even though performance varies during K-fold cross-validation experiments. The confusion matrix of Scheme 1's top model is shown in Figure 6. The program accurately identified 151 conjunctival pictures that showed patients with anemia. Moreover, true negative (TN) shows that 127 conjunctival pictures are both normal and accurately predicted to be normal by the model. Additionally, it is known from the matrix confusion that the model falsely forecasts as normal up to 26 anemic conjunctival images (False Negative (FN)) and just one normal conjunctival image (False Positive (FP)) as anemia. The confusion matrix evaluation's findings reveal how well the model can identify anemia. The model's accuracy is noteworthy at about 91.1%, meaning that roughly 91.1% of the conjunctival pictures are correctly anticipated.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

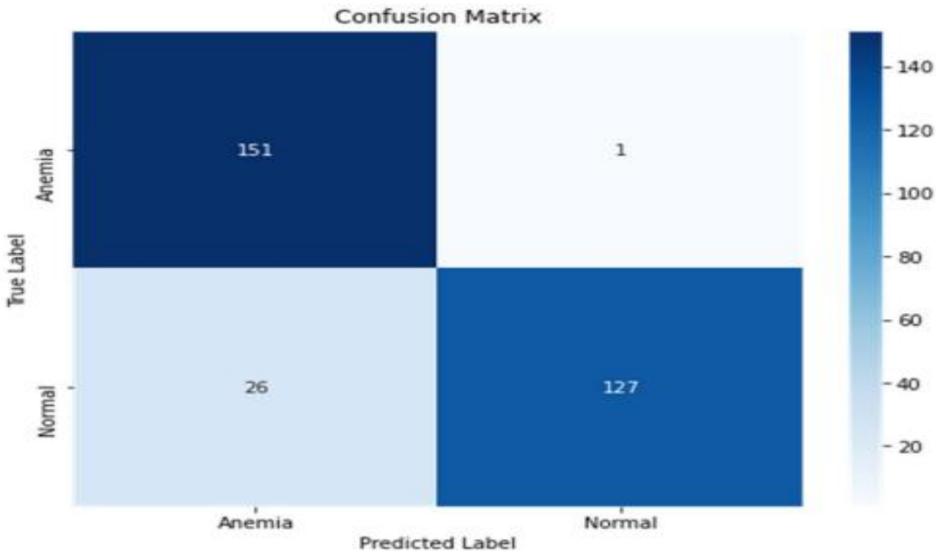

FIGURE 6. The confusion matrix in Scheme 1's best conjunctival image classification model.

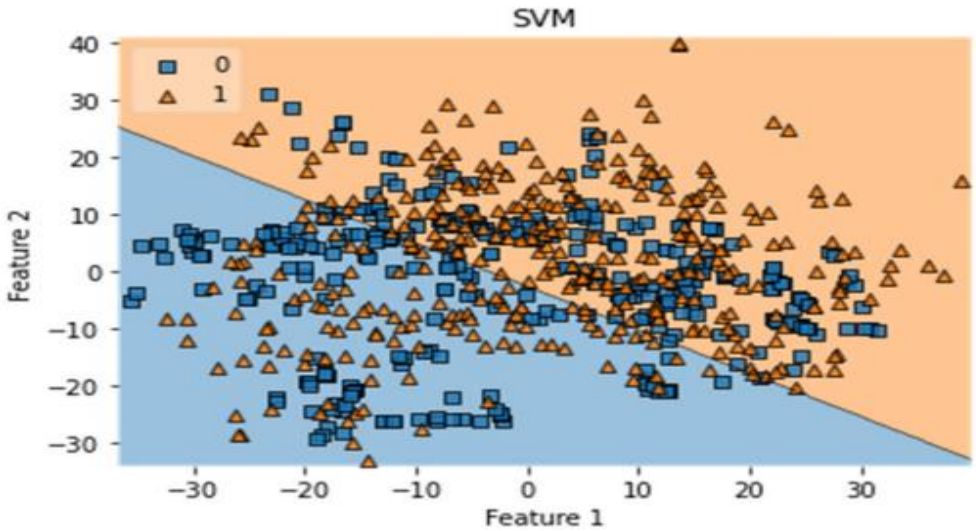

FIGURE 7. Visualization of the SVM algorithm process in conjunction with PCA to identify conjunctival images in Scheme 1.

After that, the sensitivity value, or true positive rate, is approximately 85.3%, indicating that the model does exceptionally well in detecting cases of anemia that are truly positive. On the other hand, a specificity of about 99.2% suggests that the model does an excellent job at forecasting typical cases that are actually negative. The use of Principal Component Analysis (PCA) in SVM visualization can help comprehend how the linear kernel SVM model differentiates between anemic and normal classes in two dimensions. In order to enhance data visualization, PCA is a technique for lowering feature dimensionality. In this case, the feature dimensions were reduced to two using PCA, and the outcomes were shown. Figure 7 shows the distribution of the data from the two classes on a two-dimensional plane.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

2)CONJUNCTIVAL IMAGE CLASSIFICATION USING MobileNetV2 ALGORITHM (SCHEME 2)
The MobileNetV2 algorithm is used in the second classification to identify anemia in eye conjunctival image data. At this point, the MobileNetV2 algorithm is configured using a number of parameters.
The following configuration and parameters were used:

TABLE 2. Conjunctival image classification experiment results using the MobileNetV2 algorithm (Scheme 2).

Fold	Accuracy	Sensitivity	Specificity
Fold 1	0.8622	0.811	0.937
Fold 2	0.875	0.861	0.890
Fold 3*	0.895	0.890	0.900
Fold 4	0.856	0.808	0.922
Fold 5	0.853	0.829	0.879

*The best fold in scheme 2.

1. Learning Rate (Lr):
In this study, 0.0001 is the learning rate. One factor that controls learning rate is how many steps the method takes to get the best values for the model's weights and biases in each iteration.
2. Activation:
The activation function in this work is stochastic gradient descent, or SGD. By altering the model parameters at each iteration, SGD is an optimization algorithm frequently used in machine learning to minimize the loss function.
3. Freezing Layer:
The term "freezing layer" describes the first 70% of the MobileNetV2 layer. This means that the characteristics that particular layers have already learned will be preserved during the training phase.
4. Application of L2 Regularization:
The L2 regularization overfitting prevention method involves appending the L2 normal weights component to the loss function. By doing this, the model becomes simpler and less complicated.
5. Output Layer Activation Function (Softmax):
The output layer builds the probability distribution for each class in a multiclass classification problem using the softmax function. This facilitates the identification of the most probable prediction class for the model.
6. Loss Function (Sparse Categorical Cross Entropy):
This lose function determines the degree to which the actual label and the model prediction deviate. "Sparse" indicates that the label is given as an integer and doesn't need to be converted into one-hot encoding.
Based on the results of the second scheme execution using the MobileNetV2 approach for its classification, the model tested best on the third crossvalidation (Table 2). The precision of the model's classification of all the data is demonstrated by its 0.895 accuracy score. The model has a sensitivity of 0.89, which indicates how well it can detect real anemic cases. Model specificity measures how well the model detects actual normal events; it should be 0.9.
The results of Scheme 2, which uses the MobileNetV2 algorithm for anemia detection categorization, are further explained by the confusion matrix (Figure 8). These are the results of Scheme 2's confusion matrix. As such, 135 photos were correctly classified as belonging to the "Anemia" class. It shows that 135 pictures that depicted cases of anemia were properly identified by the model. TP stands for "True Positive." A total of fifteen photos that belong to the "Anemia" class are incorrectly predicted as "Normal." According to this, the model failed to identify 15 instances of anemia (False Negative (FN)). 17 pictures that belong in the "normal" category are mistakenly predicted to be "anemia," nonetheless. As evidenced by here, the model produced a false positive (FP) by incorrectly classifying 17 normal events as anemia. It is acceptable to describe 138 photos as "normal,"

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

and they are appropriately depicted as such. This shows that 138 photos that did not show anemic TN cases were picked up by the model (True Negative).

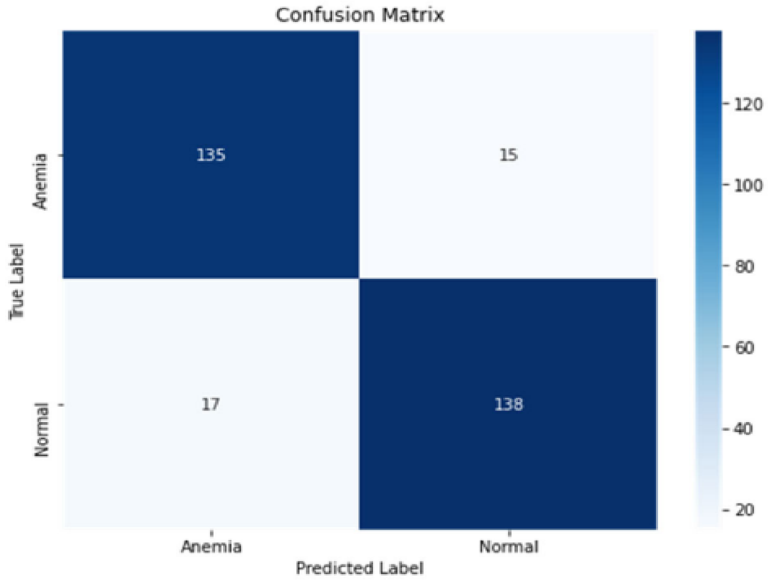

FIGURE 8. The confusion matrix in Scheme 2's best conjunctival image classification model.

3) CONJUNCTIVAL IMAGE CLASSIFICATION USING THE SVM+MobileNetV2 ALGORITHM (SCHEME 3)

A combination of the SVM and MobileNetV2 algorithms is used in the third strategy in this investigation. In this approach, MobileNetV2 is used as the first feature extractor, and the SVM algorithm is used for the subsequent classification. To put it another way, the SVM model will take the MobileNetV2 features as input. It is important to remember that during this phase, MobileNetV2's last layer—possibly the class defining layer—before the flattening layer will be frozen. In order to guarantee that the MobileNetV2 features stay constant during the SVM training procedure, this is done. The output layer is $7 \times 7 \times 1280$ in size before the flattening layer. In order to distinguish between "anemia" and "normal" circumstances, the SVM algorithm will also analyze the feature data from this frozen layer. Similar to the previous plan, the training and testing phases will be conducted, and model performance will be assessed using K-fold validation. The scheme's end results will demonstrate how the categorization of anemia detection in the eye conjunctival image dataset is impacted by the partnership of SVM and MobileNetV2. The model will be evaluated using metrics like accuracy, sensitivity, and specificity to determine how well it can differentiate between cases of anemia and normal ones. Results from the third scheme (Table 3), which combines the MobileNetV2 and SVM algorithms, show how well the model performed in the first cross-validation test. The model's ability to accurately classify all data is demonstrated by its accuracy of 0.9437. The model's ability to identify actual cases of anemia is indicated by its sensitivity of 0.915. The model's ability to identify actual normal cases is demonstrated by its specificity of 0.9718. In this system, the SVM algorithm's ideal C parameter is 1. More information regarding the third scheme's classification results is given in this confusion matrix (Figure 9). In this case, the algorithm properly identifies 143 out of 152 anemic patients as having anemia (true positive), however it mistakenly classifies 9 patients as normal even though they actually have anemia (false negative). Of the 153 people who were truly normal, 140 were diagnosed as such by the model (true negative); however, 13 were mistakenly labeled as anemic when they were actually normal (false positive).

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

These results provide insight into how well the model distinguishes between patients who are anemic and those who are not. In the subsequent step, create a PCA (Principal Component Analysis) plot for data visualization akin to Scheme 1 after completing Scheme 3, namely "SVM MobileNetV2."

TABLE 3. Conjunctival image classification experiment results using the MobileNetV2+SVM algorithm (Scheme 3).

Fold	Accuracy	Sensitivity	Specificity
Fold 1	0,9437	0,9155	0,9718
Fold 2	0,9366	0,9577	0,9155
Fold 3*	0,8944	0,8732	0,9155
Fold 4	0,9225	0,9296	0,9155
Fold 5	0,9220	0,9571	0,8873

*The best fold in scheme 3.

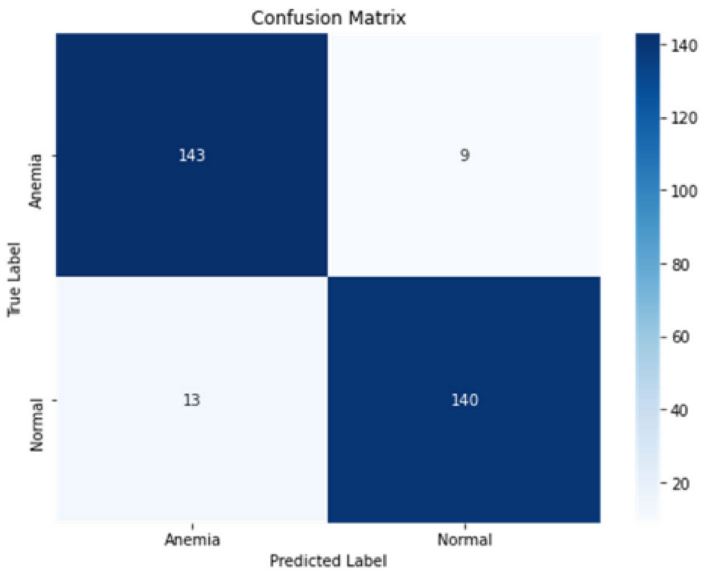

FIGURE 9. The confusion matrix is in Scheme 3's best conjunctival image classification model.

Following SVM and MobileNetV2 classification, this PCA plot will demonstrate how data from two different classes—patients Patients—both healthy and those suffering from anemia—are scattered around a two-dimensional area

TABLE 4. Ablation study of conjunctival image classification experiment results using the MobileNetV2+SVM algorithm.

Scheme	With SMOTE+TOMEK			Without SMOTE+TOMEK		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
SVM (Scheme 1)	91	82	99	78	90	29
MOBILENETV2 (Scheme 2)	89	89	90	82	82	80
MOBILENETV2+SVM (Scheme 3) *	93	91	94	83	93	42

*The best model.

Scheme 3 uses PCA visualization to display the distribution of ocular conjunctival image data in two dimensions after classification using a combination of SVM and MobileNetV2 (Figure 10) as shown below.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

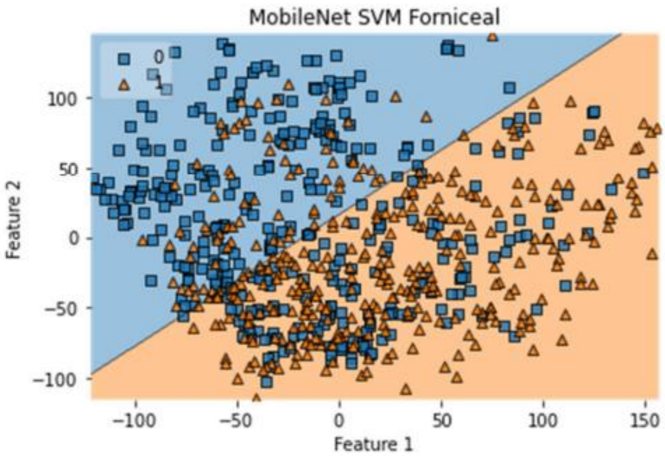

FIGURE 10. **Visualization of the SVM algorithm process in conjunction with PCA to identify conjunctival images in Scheme 3.**

This animation demonstrates how the proposed approach can distinguish between images in a two-dimensional plane that display symptoms of anemia and normal images. Each point in the PCA plot represents an image of the ocular conjunctiva from the dataset. When SVM and MobileNetV2 are used for classification, the point's location represents the image's characteristics. The model can give exact classifications if there is a distinct division between the points from the anemic and normal classes on the plot.

B. DISCUSSION

At this stage, the best test results from each scheme in this study were compared in order to assess each scheme's performance and effectiveness in identifying anemia. The results of the examination are summarized in detail in Table 4, which also rates the specificity, sensitivity, and precision of each model.
The results of the evaluation show that the SVM and MobileNetV2 models in Scheme 3 are the most successful. This demonstrates that the model developed in Scheme 3 performs the best when compared to models from Schemes 1 and 2. When compared to models in other models, the Scheme 3 model has the highest values for the accuracy, sensitivity, and specificity metrics. plans, illustrating this benefit.

TABLE 5. **Comparison anemia classification results with related work.**

Method	Accuracy	Sensitivity	Specificity
Method [13]	82	49	90
Method [30] (based vessels)	-	77	-
Method [30] (based sclerae)	-	73	-
Proposed Method	93	91	94

SVM Scheme: The first schema employs a linear kernel and the SVM algorithm. The evaluation's findings indicate that the SVM model has a 91% accuracy rate, an 82% sensitivity rate, and a 99% specificity rate. Despite having a somewhat lower sensitivity, the SVM model's high specificity shows that it is highly effective in classifying normal situations. It is still possible to enhance the capacity to identify anemic instances, nevertheless.
MobileNetV2 Scheme: The second strategy makes use of the MobileNetV2 algorithm.The evaluation's findings

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

indicate that the MobileNetV2 model has a 90% specificity, 89% sensitivity, and 89% accuracy. Although the accuracy of this model can still be increased, the sensitivity and specificity are balanced.

SVM + MobileNetV2 strategy: This third strategy combines SVM and MobileNetV2. According to the evaluation results, this model performs the best. There are three possible results: 94% specificity, 91% sensitivity, and 93% accuracy. Overall, this model can yield reasonable classification results and strikes a decent balance between identifying anemia and normal instances. According to the evaluation results, the MobileNetV2 + SVM scheme performs the best when it comes to identifying anemia in eye conjunctival images. Since this model can combine the benefits of SVM for data classification with MobileNetV2 for feature extraction, it can be said that this scheme is one that is suggested and encouraged for additional research. Ablation studies are crucial [80], and Table 4's comparison of the three schemes is considered an ablation study as well because it shows that the dependency between SVM and MobileNetV2 when the two techniques are used independently and together, the Enhanced detection performance is the end outcome. Significant results are also obtained from data augmentation based on SMOTE and data noise removal using TOMEK. Because SMOTE and TOMEK balance the dataset, Table 4 shows that even with the combination of MobileNetV2 and SVM, the detection performance may still be greatly enhanced. The well-known baselines in the field of the anemia detection method should be followed in this instance in addition to comparative ablation investigations [81]. The most crucial metrics for medical classification are accuracy, specificity, and specificity—specificity in particular. This should be emphasized once more. These findings, as shown in Table 5, are far superior to those of previous research [13] and [30]. Similar research aiming to classify anemia also makes use of the same dataset, Eyes defy-anemia. Another factor influencing this is the classification's emphasis on the palpebral region. This approach is still better than the way [13], which both make use of MobileNetV2, even when you compare the outcomes of the two systems. The sensitivity approach in particular [13] is low. In an imbalanced dataset, the positive class turns into a minor class, which needs to be underlined. Meanwhile, because it has to do with spotting positive data, sensitivity is important in the medical field. On the other hand, accuracy pertains more to general results; hence, outcomes that are unbalanced in the dataset are not suitable for use as references. Additionally, the suggested method is no better than an alternative approach to the procedure [30] that makes use of arteries and sclerae. It has been discovered that classification performance is also influenced by the appropriate methodology.

CONCLUSION

According to the findings of the conducted research, the optimal method for non-invasively detecting anemia from conjunctival pictures of the eye is to combine MobileNetV2 and SVM (SVM + MobileNetV2). In contrast to other methods, this one yields the best performance. The accuracy of the SVM + MobileNetV2 model is 93%, indicating a very strong overall capacity to categorize patients accurately. With a sensitivity of 91%, the model can reliably detect anemic individuals, lowering the possibility of a false positive. A specificity of 94%, however, shows that the model can effectively identify healthy patients. In a clinical setting, these findings are highly encouraging since they offer a simpler and non-invasive method of identifying anemia. Because SVM + MobileNetV2 allows for early detection and more efficient treatment of individuals who may have anemia, it has the potential to have a positive impact on healthcare. Even though other schemes do well as well, the SVM + MobileNetV2 combination is a great option for the non-invasive detection of anemia because it strikes a compromise between accuracy, sensitivity, and specificity. ocular conjunctival pictures.

ACKNOWLEDGEMENT

The successful completion of any task would be incomplete without a proper guidance and support. Combination of these factors acts like backbone to our TERM PAPER titled “Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI”. We are deeply grateful to our guide, Ms. S J R K PADMINIVALLI V , for her unwavering guidance, insightful feedback, and encouragement. Her expertise and dedication have been instrumental in shaping the direction and quality of this research. We would like to express our sincere gratitude to Mr. K. Siva Kumar In-charge for our term paper. His expertise, guidance, and support were instrumental in the successful completion of this research.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

REFERENCES

[1] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, ‘‘Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis,’’ Electronics, vol. 9, no. 3, p. 427, Mar. 2020, doi: 10.3390/electronics9030427.
[2] A. D’Alessandro, G. Liumbruno, G. Grazzini, and L. Zolla, ‘‘Red blood cell storage: The story so far,’’ Blood Transfus., vol. 8, no. 2, p. 82, 2010, doi: 10.2450/2009.0122-09.
[3] C. M. Chaparro and P. S. Suchdev, ‘‘Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries,’’ Ann. New York Acad. Sci., vol. 1450, no. 1, pp. 15–31, Aug. 2019, doi: 10.1111/nyas.14092.
[4] P. Paul, P. Chouhan, and A. Zaveri, ‘‘Impact of child marriage on nutritional status and anaemia of children under 5 years of age: Empirical evidence from India,’’ Public Health, vol. 177, pp. 95–101, Dec. 2019, doi: 10.1016/j.puhe.2019.08.008.
[5] A. Kumar, E. Sharma, A. Marley, M. A. Samaan, and M. J. Brookes, ‘‘Iron deficiency anaemia: Pathophysiology, assessment, practical management,’’ BMJ Open Gastroenterol., vol. 9, no. 1, Jan. 2022, Art. no. e000759, doi: 10.1136/bmjgast-2021-000759.
[6] P. W. Marks, ‘‘Anemia: Clinical approach,’’ in Concise Guide to Hematology. Cham, Switzerland: Springer, 2019, pp. 21–27.
[7] J. Dybas, F. C. Alcicek, A. Wajda, M. Kaczmarska, A. Zimna, K. Bulat, A. Blat, T. Stepanenko, T. Mohaissen, E. Szczesny-Malysiak, D. Perez-Guaita, B. R. Wood, and K. M. Marzec, ‘‘Trends in biomedical analysis of red blood cells—Raman spectroscopy against other spectroscopic, microscopic and classical techniques,’’ TrAC Trends Anal. Chem., vol. 146, Jan. 2022, Art. no. 116481, doi: 10.1016/j.trac.2021.116481.
[8] V. Jain, M. G. Munro, and H. O. D. Critchley, ‘‘Contemporary evaluation of women and girls with abnormal uterine bleeding: FIGO systems 1 and 2,’’ Int. J. Gynecol. Obstetrics, vol. 162, no. S2, pp. 29–42, Aug. 2023, doi: 10.1002/ijgo.14946.
[9] L. L. Mendes, A. M. B. V. D. Costa, M. D. P. Gonçalves, and M. F. Beirão, ‘‘Reticulocyte count from buffy coats: Medular recovery after blood donation,’’ Int. J. Health Sci., vol. 2, no. 58, pp. 2–13, Sep. 2022, doi: 10.22533/at.ed.1592582229092.
[10] H. K. Abdul-Hussein, H. S. Al-Mammori, and M. K. Hassan, ‘‘Evaluation of the expression of red blood cell CD36, interleukin-6 and interleukin8 in sickle cell anemia pediatric patients,’’ Cytokine, vol. 143, Jul. 2021, Art. no. 155534, doi: 10.1016/j.cyto.2021.155534.
[11] P.-H. Yu, M.-Y. Lin, Y.-W. Chiu, J.-J. Lee, S.-J. Hwang, C.-C. Hung, and H.-C. Chen, ‘‘Low serum iron is associated with anemia in CKD stage 1–4 patients with normal transferrin saturations,’’ Sci. Rep., vol. 11, no. 1, pp. 1–10, Apr. 2021, doi: 10.1038/s41598-021-87401-w.
[12] M. D. Cappellini, K. M. Musallam, and A. T. Taher, ‘‘Iron deficiency anaemia revisited,’’ J. Internal Med., vol. 287, no. 2, pp. 153–170, Feb. 2020, doi: 10.1111/joim.13004.
[13] G. Dimauro, M. E. Griseta, M. G. Camporeale, F. Clemente, A. Guarini, and R. Maglietta, ‘‘An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset,’’ Artif. Intell. Med., vol. 136, Feb. 2023, Art. no. 102477, doi: 10.1016/j.artmed.2022.102477.
[14] N. Kobayashi, A. Yoshino, M. Ishikawa, and S. Homma, ‘‘Anemia examination using a hyperspectral camera in telecare system,’’ in Proc. IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech), Mar. 2021, pp. 475–476, doi: 10.1109/LifeTech52111.2021.9391912.
[15] M. Jaiswal, A. Srivastava, and T. J. Siddiqui, ‘‘Machine learning algorithms for anemia disease prediction,’’ in Recent Trends in Communication, Computing, and Electronics(Lecture Notes in Electrical Engineering), vol. 524. Singapore: Springer, 2019, pp. 463–469.
[16] A. Dixit, R. Jha, R. Mishra, and S. Vhatkar, ‘‘Prediction of anemia disease using machine learning algorithms,’’ in Intelligent Computing and Networking. Singapore: Springer, 2023, pp. 229–238.
[17] S. De and B. Chakraborty, ‘‘Case-based reasoning (CBR)-based anemia severity detection system (ASDS) using machine learning algorithm,’’ in Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl. Singapore: Springer, 2021, pp. 621–632.
[18] M. A. Araaf, K. Nugroho, and D. R. I. M. Setiadi, ‘‘Comprehensive analysis and classification of skin diseases based on image texture features using K-nearest neighbors algorithm,’’ J. Comput. Theories Appl., vol. 1, no. 1, pp. 31–40, Sep. 2023, doi: 10.33633/jcta.v1i1.9185.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[19] F. Mustofa, A. N. Safriandono, A. R. Muslikh, and D. R. I. M. Setiadi, ‘‘Dataset and feature analysis for diabetes mellitus classification using random forest,’’ J. Comput. Theories Appl., vol. 1, no. 1, pp. 41–48, Sep. 2023, doi: 10.33633/jcta.v1i1.9190.
[20] L. Ji, S. Wu, and X. Gu, ‘‘A facial expression recognition algorithm incorporating SVM and explainable residual neural network,’’ Signal, Image Video Process., vol. 17, no. 8, pp. 4245–4254, Nov. 2023, doi: 10.1007/s11760-023-02657-1.
[21] X. Yang, Z. Hua, L. Zhang, X. Fan, F. Zhang, Q. Ye, and L. Fu, ‘‘Preferred vector machine for forest fire detection,’’ Pattern Recognit., vol. 143, Nov. 2023, Art. no. 109722, doi: 10.1016/j.patcog.2023.109722.
[22] P. Tavana, M. Akraminia, A. Koochari, and A. Bagherifard, ‘‘Classification of spinal curvature types using radiography images: deep learning versus classical methods,’’ Artif. Intell. Rev., vol. 56, no. 11, pp. 13259–13291, 2023, doi: 10.1007/s10462-023-10480-w.
[23] S. Wang, B. Liu, Y.-L. Wang, Y. Hu, J. Liu, X.-D. He, J. Yuan, and Q. Wu, ‘‘Machine learning-based human motion recognition via wearable plastic fiber sensing system,’’ IEEE Internet Things J., vol. 10, no. 20, pp. 17893–17904, Oct. 2023, doi: 10.1109/JIOT.2023. 3277829.
[24] M. S. Sunarjo, H.-S. Gan, and D. R. I. M. Setiadi, ‘‘High-performance convolutional neural network model to identify COVID-19 in medical images,’’ J. Comput. Theories Appl., vol. 1, no. 1, pp. 19–30, Aug. 2023, doi: 10.33633/jcta.v1i1.8936.
[25] R. N. Wessner, R. Frozza, D. Duarte da Silva Bagatini, and R. F. Molz, ‘‘Recognition of weeds in corn crops: System with convolutional neural networks,’’ J. Agricult. Food Res., vol. 14, Dec. 2023, Art. no. 100669, doi: 10.1016/j.jafr.2023.100669.
[26] M. H. Rahman, M. K. A. Jannat, M. S. Islam, G. Grossi, S. Bursic, and M. Aktaruzzaman, ‘‘Real-time face mask position recognition system based on MobileNet model,’’ Smart Health, vol. 28, Jun. 2023, Art. no. 100382, doi: 10.1016/j.smhl.2023.100382.
[27] H. T. Adityawan, O. Farroq, S. Santosa, H. M. M. Islam, M. K. Sarker, and D. R. I. M. Setiadi, ‘‘Butterflies recognition using enhanced transfer learning and data augmentation,’’ J. Comput. Theories Appl., vol. 1, no. 2, pp. 115–128, Nov. 2023, doi: 10.33633/jcta.v1i2.9443.
[28] H. Shrestha, S. C. B. Jaganathan, C. Dhasarathan, and K. Suriyan, ‘‘Detection and classification of dermatoscopic images using segmentation and transfer learning,’’ Multimedia Tools Appl., vol. 82, no. 15, pp. 23817–23831, Jun. 2023, doi: 10.1007/s11042-023-14752-z.
[29] X. Wu, Z. Luo, and H. Xu, ‘‘Recognition of pear leaf disease under complex background based on DBPNet and modified mobilenetV2,’’ IET Image Process., vol. 17, no. 10, pp. 3055–3067, Aug. 2023, doi: 10.1049/ipr2.12855.
[30] G. Dimauro, M. G. Camporeale, A. Dipalma, A. Guarini, and R. Maglietta, ‘‘Anaemia detection based on sclera and blood vessel colour estimation,’’ Biomed. Signal Process. Control, vol. 81, Mar. 2023, Art. no. 104489, doi: 10.1016/j.bspc.2022.104489.
[31] P. Soltanzadeh and M. Hashemzadeh, ‘‘RCSMOTE: Range-controlled synthetic minority over-sampling technique for handling the class imbalance problem,’’ Inf. Sci., vol. 542, pp. 92–111, Jan. 2021, doi: 10.1016/j.ins.2020.07.014.
[32] Q. Ning, X. Zhao, and Z. Ma, ‘‘A novel method for identification of glutarylation sites combining borderline-SMOTE with tomek links technique in imbalanced data,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 19, no. 5, pp. 2632–2641, Sep. 2022, doi: 10.1109/TCBB.2021. 3095482.
[33] P. Appiahene, J. W. Asare, E. T. Donkoh, G. Dimauro, and R. Maglietta, ‘‘Detection of iron deficiency anemia by medical images: A comparative study of machine learning algorithms,’’ BioData Mining, vol. 16, no. 1, pp. 1–20, 2023. [Online]. Available: https://doi.org/10.1186/s13040-023- 00319-z
[34] R. G. Mannino, D. R. Myers, E. A. Tyburski, C. Caruso, J. Boudreaux, T. Leong, G. D. Clifford, and W. A. Lam, ‘‘Smartphone app for noninvasive detection of anemia using only patient-sourced photos,’’ Nature Commun., vol. 9, no. 1, p. 4924, Dec. 2018, doi: 10.1038/s41467-018- 07262-2.
[35] G. Dimauro and L. Simone, ‘‘Novel biased normalized cuts approach for the automatic segmentation of the conjunctiva,’’ Electronics, vol. 9, no. 6, p. 997, Jun. 2020, doi: 10.3390/electronics9060997.
[36] E. F. Swana, W. Doorsamy, and P. Bokoro, ‘‘Tomek link and SMOTE approaches for machine fault classification with an imbalanced dataset,’’ Sensors, vol. 22, no. 9, p. 3246, Apr. 2022, doi: 10.3390/s22093246.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[37] S. Wang, Y. Dai, J. Shen, and J. Xuan, ''Research on expansion and classification of imbalanced data based on SMOTE algorithm,'' Sci. Rep., vol. 11, no. 1, pp. 1–11, Dec. 2021, doi: 10.1038/s41598-021-03430-5.
[38] S. Demir and E. K. Sahin, ''Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data,'' Soil Dyn. Earthq. Eng., vol. 154, Mar. 2022, Art. no. 107130, doi: 10.1016/j.soildyn.2021.107130.
[39] S. Demir and E. K. Sahin, ''An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost,'' Neural Comput. Appl., vol. 35, no. 4, pp. 3173–3190, 2023, doi: 10.1007/S00521-022-07856-4.
[40] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, ''A review of medical image data augmentation techniques for deep learning applications,'' J. Med. Imag. Radiat. Oncol., vol. 65, no. 5, pp. 545–563, Aug. 2021, doi: 10.1111/1754-9485.13261.
[41] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliunas, ¯ ''Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning,'' Expert Syst., vol. 38, no. 7, p. e12746, Nov. 2021, doi: 10.1111/exsy.12746.
[42] K. M. Hosny, M. A. Kassem, and M. M. Foaud, ''Classification of skin lesions using transfer learning and augmentation with alex-net,'' PLoS ONE, vol. 14, no. 5, May 2019, Art. no. e0217293, doi: 10.1371/journal.pone.0217293.
[43] M. Liu, W. Yang, X. Zhu, J. Chen, X. Chen, L. Yang, and E. H. Helmer, ''An improved flexible spatiotemporal DAta fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series,'' Remote Sens. Environ., vol. 227, pp. 74–89, Jun. 2019, doi: 10.1016/j.rse.2019.03.012.
[44] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, ''Secure, privacy-preserving and federated machine learning in medical imaging,'' Nature Mach. Intell., vol. 2, no. 6, pp. 305–311, Jun. 2020, doi: 10.1038/s42256-020-0186-1.
[45] N. Moshkov, B. Mathe, A. Kertesz-Farkas, R. Hollandi, and P. Horvath, ''Test-time augmentation for deep learning-based cell segmentation on microscopy images,'' Sci. Rep., vol. 10, no. 1, pp. 1–7, Mar. 2020, doi: 10.1038/s41598-020-61808-3.
[46] R. Hao, K. Namdar, L. Liu, M. A. Haider, and F. Khalvati, ''A comprehensive study of data augmentation strategies for prostate cancer detection in diffusion-weighted MRI using convolutional neural networks,'' J. Digit. Imag., vol. 34, no. 4, pp. 862–876, 2021, doi: 10.1007/S10278-021- 00478-7.
[47] A. Timofeev, A. Fadeeva, A. Afonin, C. Musat, and A. Maksai, ''DSS: Synthesizing long digital ink using data augmentation, style encoding and split generation,'' in Proc. Int. Conf. Document Anal. Recognit., 2023, pp. 217–235, doi: 10.1007/978-3-031-41685-9_14.
[48] C. I. Nwoye, T. Yu, C. Gonzalez, B. Seeliger, P. Mascagni, D. Mutter, J. Marescaux, and N. Padoy, ''Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos,'' Med. Image Anal., vol. 78, May 2022, Art. no. 102433, doi: 10.1016/j.media.2022.102433.
[49] N. V. Sharma and N. S. Yadav, ''An optimal intrusion detection system using recursive feature elimination and ensemble of classifiers,'' Microprocessors Microsyst., vol. 85, Sep. 2021, Art. no. 104293, doi: 10.1016/j.micpro.2021.104293.
[50] R. Amanda and E. S. Negara, ''Analysis and implementation machine learning for Youtube data classification by comparing the performance of classification algorithms,'' J. Online Inform., vol. 5, no. 1, pp. 61–72, 2020. [Online]. Available: https://doi.org/10.15575/join.v5i1.505
[51] V. Sampath, I. Maurtua, J. J. A. Martín, A. Iriondo, I. Lluvia, and G. Aizpurua, ''Intraclass image augmentation for defect detection using generative adversarial neural networks,'' Sensors, vol. 23, no. 4, p. 1861, 2023, doi: 10.3390/S23041861.
[52] A. Mumuni and F. Mumuni, ''Data augmentation: A comprehensive survey of modern approaches,'' Array, vol. 16, Dec. 2022, Art. no. 100258, doi: 10.1016/j.array.2022.100258.
[53] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''SMOTE: Synthetic minority over-sampling technique,'' J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[54] R. M. Pereira, Y. M. G. Costa, and C. N. Silla Jr., ‘‘MLTL: A multi-label approach for the tomek link undersampling algorithm,’’ Neurocomputing, vol. 383, pp. 95–105, Mar. 2020, doi: 10.1016/j.neucom.2019.11.076.
[55] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, ‘‘PTB-XL, a large publicly available electrocardiography dataset,’’ Sci. Data, vol. 7, no. 1, pp. 1–15, May 2020, doi: 10.1038/s41597-020-0495-6.
[56] S. Bauskar, P. Jain, and M. Gyanchandani, ‘‘A noninvasive computerized technique to detect anemia using images of eye conjunctiva,’’ Pattern Recognit. Image Anal., vol. 29, no. 3, pp. 438–446, 2019, doi: 10.1134/S1054661819030027.
[57] G. Battineni, N. Chintalapudi, and F. Amenta, ‘‘Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM),’’ Informat. Med. Unlocked, vol. 16, Jan. 2019, Art. no. 100200, doi: 10.1016/j.imu.2019.100200.
[58] H. Al Azies and G. Anuraga, ‘‘Classification of underdeveloped areas in Indonesia using the SVM and k-NN algorithms,’’ Jurnal ILMU DASAR, vol. 22, no. 1, p. 31, Jan. 2021, doi: 10.19184/jid.v22i1.16928.
[59] N. Shafaf and H. Malek, ‘‘Applications of machine learning approaches in emergency medicine; a review article,’’ Arch. Acad. Emergency Med., vol. 7, no. 1, p. e34, 2019, doi: 10.22037/aaem.v7i1.410.
[60] H. Al Azies, D. Trishnanti, and E. P. H. Mustikawati, ‘‘Comparison of kernel support vector machine (SVM) in classification of human development index (HDI),’’ IPTEK J. Proc. Ser., vol. 0, no. 6, pp. 53–57, 2019. [Online]. Available: https://doi.org/10.1080/01621459.2023.2197686
[61] S. Rahman, F. M. Javed, M. Shamrat, Z. Tasnim, J. Roy, and S. A. Hossain, ‘‘A comparative study on liver disease prediction using supervised machine learning algorithms,’’ Int. J. Sci. Technol. Res., vol. 8, no. 11, pp. 419–422, 2019. [Online]. Available: http://www.ijstr.org
[62] F. Anowar, S. Sadaoui, and B. Selim, ‘‘Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE),’’ Comput. Sci. Rev., vol. 40, May 2021, Art. no. 100378, doi: 10.1016/j.cosrev.2021.100378.
[63] J. Bharadiya and J. P. Bharadiya, ‘‘A tutorial on principal component analysis for dimensionality reduction in machine learning,’’ Int. J. Innov. Res. Sci. Eng. Technol., vol. 8, no. 5, pp. 2028–2032, 2023, doi: 10.5281/zenodo.8002436.
[64] G. H. Christa, J. Jesica, A. Anisha, and K. M. Sagayam, ‘‘CNN-based mask detection system using OpenCV and MobileNetV2,’’ in Proc. 3rd Int. Conf. Signal Process. Commun. (ICPSC), May 2021, pp. 115–119, doi: 10.1109/ICSPC51351.2021.9451688.
[65] S. B. Imanulloh, A. R. Muslikh, and D. R. I. M. Setiadi, ‘‘Plant diseases classification based leaves image using convolutional neural network,’’ J. Comput. Theories Appl., vol. 1, no. 1, pp. 1–10, Aug. 2023, doi: 10.33633/jcta.v1i1.8877.
[66] K. Dong, C. Zhou, Y. Ruan, and Y. Li, ‘‘MobileNetV2 model for image classification,’’ in Proc. 2nd Int. Conf. Inf. Technol. Comput. Appl. (ITCA), Dec. 2020, pp. 476–480, doi: 10.1109/ITCA52113.2020.00106.
[67] P. Arafin, A. Issa, and A. H. M. M. Billah, ‘‘Performance comparison of multiple convolutional neural networks for concrete defects classification,’’ Sensors, vol. 22, no. 22, p. 8714, Nov. 2022, doi: 10.3390/s22228714.
[68] A. Agrawal. (2021). Detecting Anemia From Retinal Images Using Deep Learning. [Online]. Available: http://localhost:8080/ xmlui/handle/10263/7292
[69] T. F. Mahdi, H. G. Daway, and J. Jouda, ‘‘White blood cell detection and classification using transfer Densenet201 and Mobilenetv2 learning models,’’ AIP Conf., vol. 2830, no. 1, p. 40013, 2023, doi: 10.1063/5.0156771.
[70] S. Taufiqurrahman, A. Handayani, B. R. Hermanto, and T. L. E. R. Mengko, ‘‘Diabetic retinopathy classification using a hybrid and efficient MobileNetV2-SVM model,’’ in Proc. IEEE REGION Conf. (TENCON), Nov. 2020, pp. 235–240, doi: 10.1109/TENCON50793.2020.9293739.
[71] A. Michele, V. Colin, and D. D. Santika, ‘‘MobileNet convolutional neural networks and support vector machines for palmprint recognition,’’ Proc. Comput. Sci., vol. 157, pp. 110–117, Jan. 2019, doi: 10.1016/j.procs.2019.08.147.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

[72] N. Mathai, Y. Chen, and J. Kirchmair, ''Validation strategies for target prediction methods,'' Briefings Bioinf., vol. 21, no. 3, pp. 791–802, May 2020, doi: 10.1093/bib/bbz026.
[73] S. Bates, T. Hastie, and R. Tibshirani, ''Cross-validation: What does it estimate and how well does it do it?'' J. Amer. Stat. Assoc., pp. 1–12, May 2023, doi: 10.1080/01621459.2023.2197686.
[74] S. Prusty, S. Patnaik, and S. K. Dash, ''SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer,'' Frontiers Nanotechnol., vol. 4, Aug. 2022, Art. no. 972421, doi: 10.3389/FNANO.2022.972421.
[75] A. A. Winoto and A. F. V. Roy, ''Model of predicting the rating of bridge conditions in Indonesia with regression and K-fold cross validation,'' Int. J. Sustain. Construction Eng. Technol., vol. 14, no. 1, pp. 249–259, Feb. 2023, doi: 10.30880/ijscet.2023.14.01.022.
[76] G. N. Ahmad, H. Fatima, S. Ullah, A. S. Saidi, and Imdadullah, ''Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV,'' IEEE Access, vol. 10, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
[77] M. Grandini, E. Bagli, and G. Visani, ''Metrics for multi-class classification: An overview,'' 2020, arXiv:2008.05756.
[78] C. M. Burlacu, A. C. Burlacu, and M. Praisler, ''Sensitivity analysis of artificial neural networks identifying JWH synthetic cannabinoids built with alternative training strategies and methods,'' Inventions, vol. 7, no. 3, p. 82, Sep. 2022, doi: 10.3390/inventions7030082.
[79] M. Jahangiri, M. Jahangiri, and M. Najafgholipour, ''The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran,'' Sci. Total Environ., vol. 728, Aug. 2020, Art. no. 138872, doi: 10.1016/j.scitotenv.2020.138872.
[80] G. Xu, M. Hu, and C. Ma, ''Secure and smart autonomous multi-robot systems for opinion spammer detection,'' Inf. Sci., vol. 576, pp. 681–693, Oct. 2021, doi: 10.1016/j.ins.2021.07.072. [81] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, ''CorrAUC: A malicious bot-IoT traffic detection method in IoT network using machine-learning techniques,'' IEEE Internet Things J., vol. 8, no. 5, pp. 3242–3254, Mar. 2021, doi: 10.1109/JIOT.2020.3002255..