Malaria Detection Using Blood Smear Images

Nishi Yadav*, Dasari Gayathri, Maddula Sai Sunamdha Harinhi

[1]School of Studies inEngineering and Technology Guru Ghasidas Vishwavidyalaya Bilaspur, India

nyadav.cse@gmail.com, amma.vsp@gmail.com, sunamdhamaddula999@gmail.com

Abstract—Malaria is one of the deadliest diseases. A perilous disease provoked by parasites that can disseminated to people from the nips of infected female mosquitoes which belongs to Anopheles genus. It is preventable and curable. However, Malaria interpretation entail close scrutiny of the blood smear images at 100x enlargement. This is followed by a freehand computing process in which adepts tally the count of Red blood cells influenced by parasites. Perception of images of malaria blood smear is a upgradeable self-activating quick fix which extricate a sea of time for medical sector plod along struggle odds with this pernicious disease. In this work, we set out to recognize from images of blood smear using deep learning methods to predict in case the sample is taken from healthy person or not. Here, we use SVM_HOG a deep learning technique to classify images in Parasitized/Uninfected images in which we use almost 19290 images of cells where it contain similar amount of both infected and uninfected images from the Kaggle database, from image we extracted the HOG Features after extracting features we feed to a classifier SVM to predict whether itis a Parasitized/Uninfected images the accuracy of our model using the data set of malaria blood smear images, we attained an accuracy of 92.69% using Linear SVM as a classifier . The results suggest that it has high accuracy on comparison with other techniques.

Keywords—Images of blood smear, Linear Support_Vector_Machine(SVM), Histogram_Oriented _gradients-HOG features, malaria parasites.

1. INTRODUCTION

Malaria a lethal disease causes because of parasites called malaria vectors and disseminated to people with the nips of infected female mosquitoes which belongs to Anopheles genus. It is avoidable and treatable. The approximate value of people died due to malaria are 409000 in 2019. Children are aged under 5 years the most exposed group affected by malaria; they reckoned about 274000 which is around 67% demises due to malaria at global in 2019.WHO African nation schleps a gratuitous high allowance of malaria worldwide load. African zone was abode to94% malaria victim or demises in 2019[1]. Ambiguous upshots such as False positive clinical outcome give rise to needless avail of antiprotozoal drugs which successively leads to stomach-ache, diarrhoea, illness, vomiting etc. whereabout false negative clinical outcome causes superfluous use of medication, subsequent confirmation, and which boost dreadful conditions of malaria [2]. So, the evolution of self-moving procedure for treatment of malaria is an enchanting probing intent for modifying each and every patient treatment and controlling. Automatic parasite spotting has large head start such as it can provide a betterand reliable treatment, especially when limited resource available, and clinical costs will decrease. Each slide of blood on microscopic probe, as well as measurable parasite perception and genus identification takes a specialist in microscopy around fifteen to thirty proceedings. In a view of 100-1000 of blood smears are scrutinized manually on an annual basis this leads to a enormous economic battle needed for diagnosis of malaria.[3] Detection of malaria in earlier stage helpful for treatment of patient and can reduce dangerous consequences. Giemsa stain on microscopic inspection of malaria parasite is glaring [4]. Other copious techniques such as expeditious diagnostic trials, reaction of polymerase chain in blood smears to identify on the presence of antigens. Even though different tests surmount in malaria identification, anyhow microscopy is ubiquitous due to inexpensive and less cumbersome and its success

rely upon pathologist prowess [5].

2. RELATED WORKS

Diaz [6] This paper is about categorization images of blood smear automatically, techniques of machine learning are used e.g., Diaz et al. blood smear pictures categorization utilizing a Support_Vector_Machine (SVM) to spot stage of their infection the contaminated red blood cells. This technique was given out good results with 94.0 percent sensitivity on a data which contains four fifty pictures. Tek [7], this paper is about computer vision- on malarial parasite identification studies eg., Tek et al. made a use of developed KNN following normalization and correcting of colour among nine blood input picture films for binary categorization. In this paper [8] they proposed a CNN architecture which was stacked which make detection of the malaria much better using procedures like five- folded cross validation on around 27558 pictures which consists of equal amount of both healthy and infected pictures of cell and they got an accuracy of 99.98 percent with a hundred percent precision and 99.9% recall. This paper [9] proposes the CNN based-model on sixteen layers parasite of malaria detection which differentiate the blood cells as contaminated or decontaminate. This framework gained 97.0 percent accuracy, a n d a higher specificity and sensitivity of than learning was upskilled with roughly 27000 images with image size forty-four x forty-four pixels. The recognition of smears of blood for plasmodium using concentrated Leishman stain on stacked images proposed by Gopa kumar et al., [10] with an architecture that is customized CNN gave a result of 97.0% sensitivity (97.0%) and specificity (98.0%). The paper [11], the malarial parasite automatic observation was explained by evaluating at patient level and thumbnails with gross 94.1% specificity ,89.7%precision, and 89.7% sensitivity to upgrade the user confidence in system findings. This paper [12] the main attention is using convolution neural network to find full count of pictures of blood smear. If malarial pathogens are available, the network is upskilled to spot them. Demonstrations specifies that the mean average precision of the gross production of the system over 0.95 when contrasted with the base truth. Additionally, the system forecasts the images containing malarial parasites as contaminated 100percent of the time. For quick prototyping, they ported the software to minimum priced microcomputer. This paper [13] suggests an artificial neural network combined with autoencoder that is stacked sparsely they used softmax classifier with 2 nodes in output layer along with they used 10- fold cross validation which proves that it work with new dataset using they got an accuracy of around 89.10% and sensitivity 93.9% and specificity of 83.1%. In this research [15], a reduction approach of a hybrid dimension suggests a algorithm which is genetic and optimize to get a proper subset of features from available data. Kernel classifiers used is Support_Vector_Machine's (SVM) classifier utilized the reduced malaria vector dataset to assess the classification performance of the experiment. In this paper [16] with the help of photo acoustics surface acoustic wave sensor (PA_SAW) malaria can be treated in earlier stage. And its sensing system categorize ordinary blood from the blood which is infected of one percent concentration is identified. In this work [17] they concentrated on study of mosquitoes which spread malaria at first, they collected mosquitos' wings as well as bodies scattering properties and observed 808 nm polarized near infrared light befit for identifying Anopheles mosquitos' wings. In this [18] research they proposed a sensor which is portable, reusable which is sensitive and available at low price it identifies hemozoin pictograms at (12.7,25.4) it mainly deals with varying hemoglobin magnetic property due to parasite of malaria. In this paper [19], they used pipeline with image size of large 5312×2988 for identification of red blood cells and adding up images of thin blood smear called RBCNET they used 2 lakh labeled cells of 965 images from around 193 patients

which give high true positive rate. In this paper they concentrated on detecting hemozoin at less concentration. At first for malaria treatment, they reviewed a lot of Raman spectroscopy methods with sample postprocessing they confirmed the Nano silver particles after disintegration of parasite and RBC cells that helps to judge quality of antimalaria drugs. In [20] [21], the authors used image data analysis for finding healthcare disease using machine learning technique.

3. MATERIALS AND METHODS

In this part we extracted features of image using HOG, or Histogram of Oriented Gradients, which is used very often for getting the features from image as data and followed by we feed up our model against SVM classifier for malaria parasite detection on malaria blood smear images.

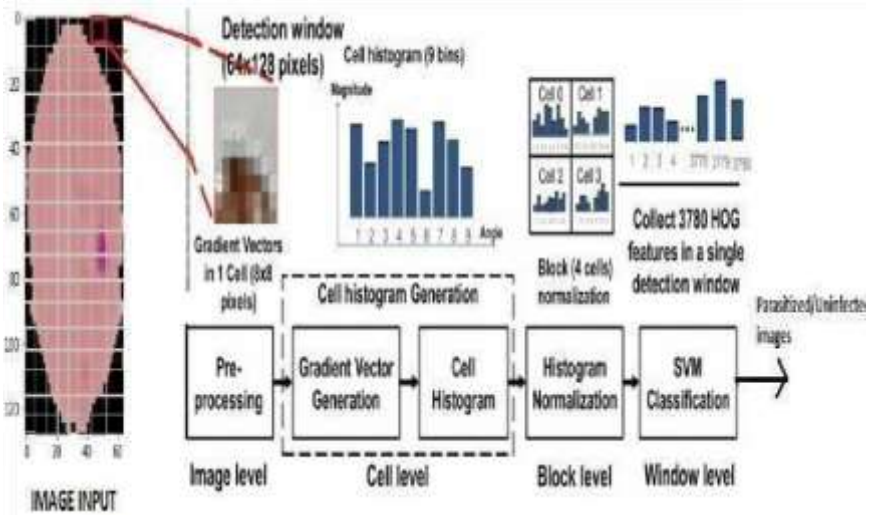

Fig-1 Hog_Svm for perception of malaria [14]

3.1.*Dataset and Preprocessing*

- **Dataset Collection**

The images used in making both the training validation datasets are a collection of 19290 blood smear images that are acquired from Kaggle datasets.

- **Data Preparation**

We created cell images folder contains all the images of the dataset and the file train.csv contain image names belonging to dataset and their corresponding labels that is Parasitized/Uninfected and set the base directory for reading images as all the images of the dataset. It is observed that our train set consists of equal samples of both the classes thus we will not face any problem due to class imbalance in the dataset. since having textual labels for our images i.e., parastized/Uninfected in the train.csv file so we have to change them to numerical labels i.e. 0 or 1. And we divide the dataset into two parts such as training and validation sets. Training set is the subset of the dataset that is used for training our model when it comes to validation set tells whether our model working upto the mark of the model after each epoch. We can observe there are images of different shapes in fig-2. It is necessary to have images in shape size before going ahead with modeling process and it is also dependent on which feature extractor tool. <Figure size

1080x1080 with 0 Axes>

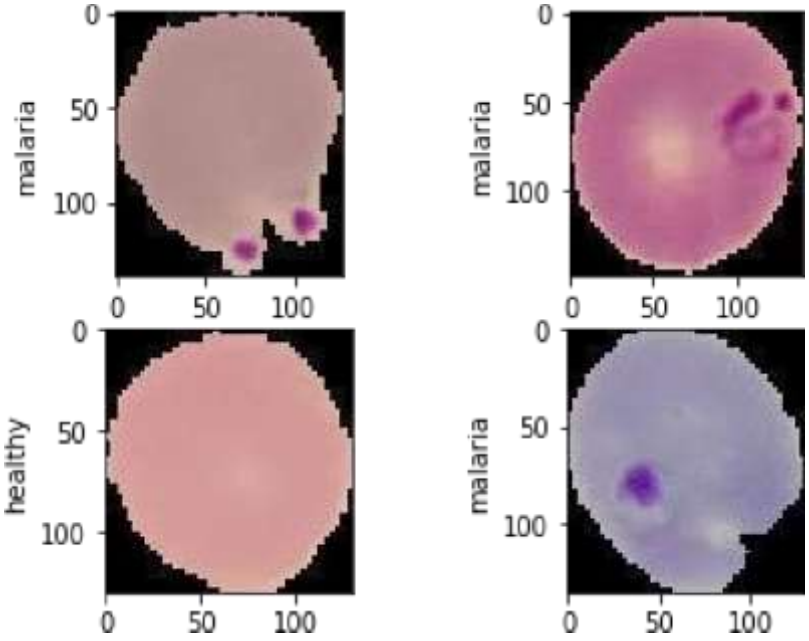

Fig-2. Images of cells

3.2. *HOG_ Features and Linear_ SVM*

Extraction of Images: Histogram of Oriented Gradients (HOG) is an identifier of a feature which is used often for feature extraction from images. In fields such as computer vision HOG entrenched. At begin Histogram_of_oriented_gradients features put to use by Navnet Dalal which highlights finding the of human. However dissimilar from other geometry features which considered integrity of an image, HOG feature, at first divides the picture into small parts known as cells and after that we calculate gradients and accumulates the histogram of gradient of the cell over each pixel. The whole blocks concatenated histograms form the HOG vectors. For extracting HOG feature primarily, the image has to be made into small pieces and for each piece we calculated gradient magnitude as well as gradient direction. We then plot Histogram using the Gradient magnitude and direction.

"The descriptor of HOG feature determines the gradient orientation contingency in the image localized portions."

- Techniques of calculating the Histogram of Oriented Gradients

From the image below which is of 128x128 we have to extract Hog features.

- Preprocess the Data

Before extracting of features from image we need to resize it to 64x128 as shown in figure 3. after that we divide the image into 8*8 and 2*2 patches in both vertical and horizontal respectively.

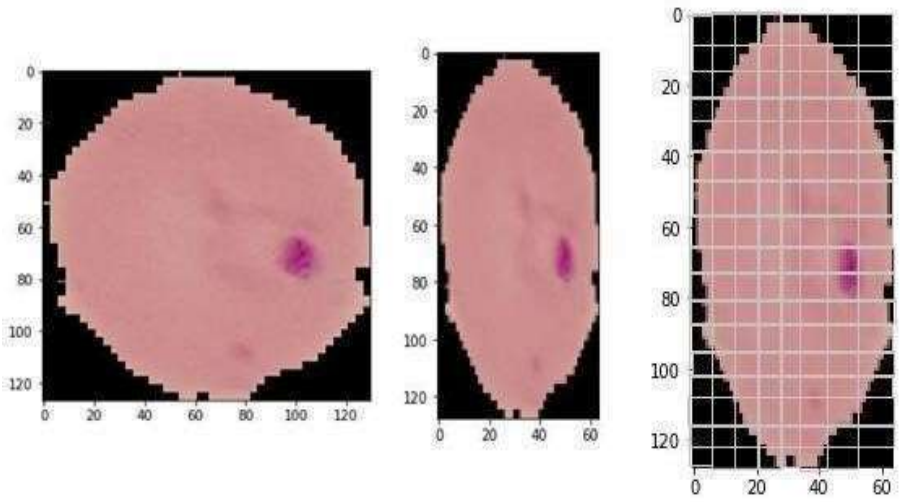

Fig-3. Preprocessed (64*128) and made(8*16) cells

Gradients Calculation in both X and Y directions: For calculation of gradient for every patch as shown in Fig 4. Variation in x and the y directions treated as gradients. From the highlighted patch of the image, we calculated the gradients. And from that we generated pixel matrix the matrix shown below is just for explanation they are not actual pixels.

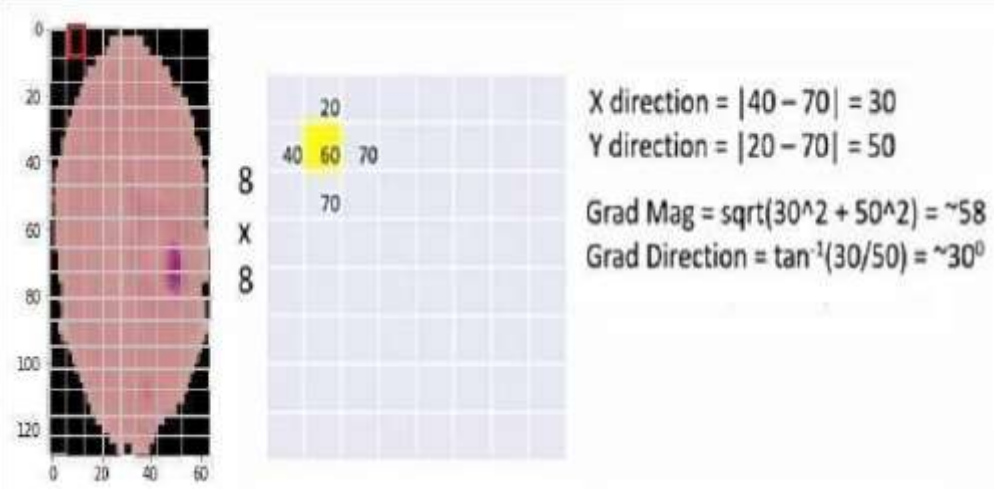

Fig-4. Measuring gradients magnitude and their direction

A histogram plot shown on fig-5 is the continuous distribution of a group of data. We took variable bins on x axis and their frequencies on y-axis. In Short Magnitude on y-axis and orientation x-axis. This is how histograms are as shown in fig 6.

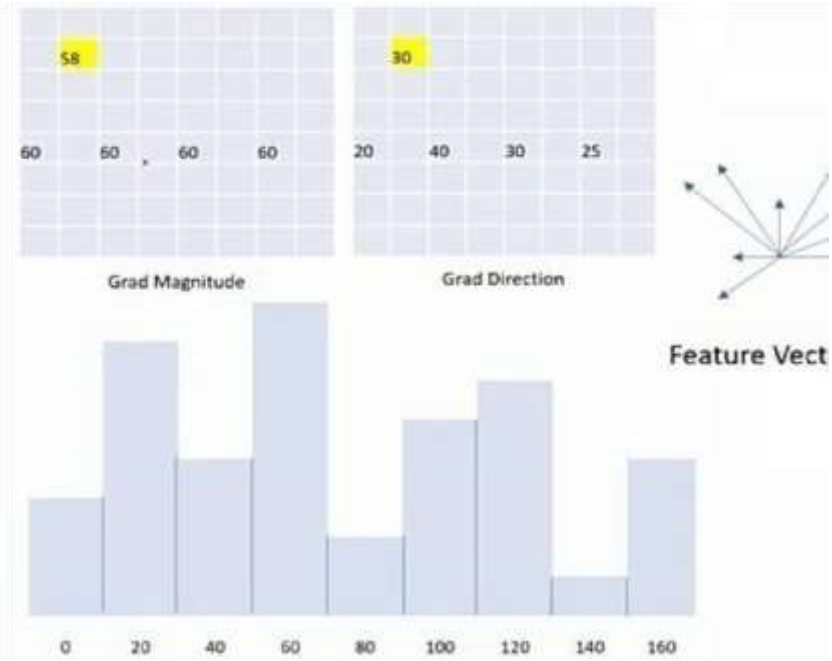

Fig-5: Histogram

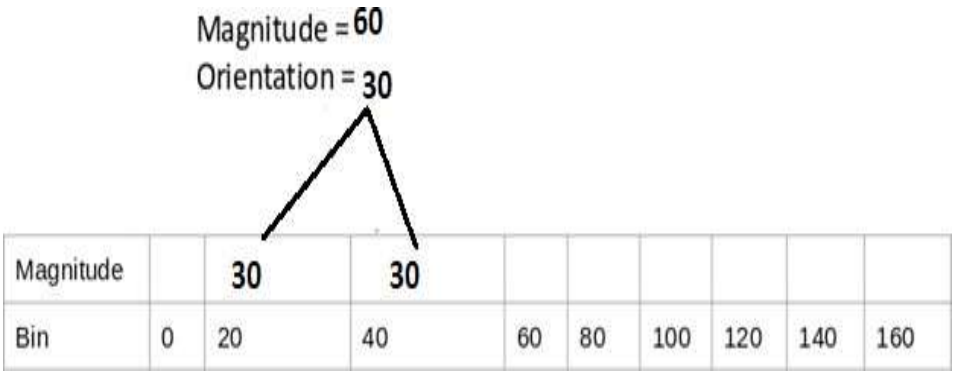

Fig-6. Measuring magnitude and orientation.

1) Calculation of Histogram of Gradients in 8×8 cells and conversion into 9×1 vector.

The single patch is divided into 8x8 matrix and from that patch we generate pixel matrix after generation we calculate gradient orientation as well as direction after which we plot a histogram with nine bins which is none another than vector of size (9*1) matrix.

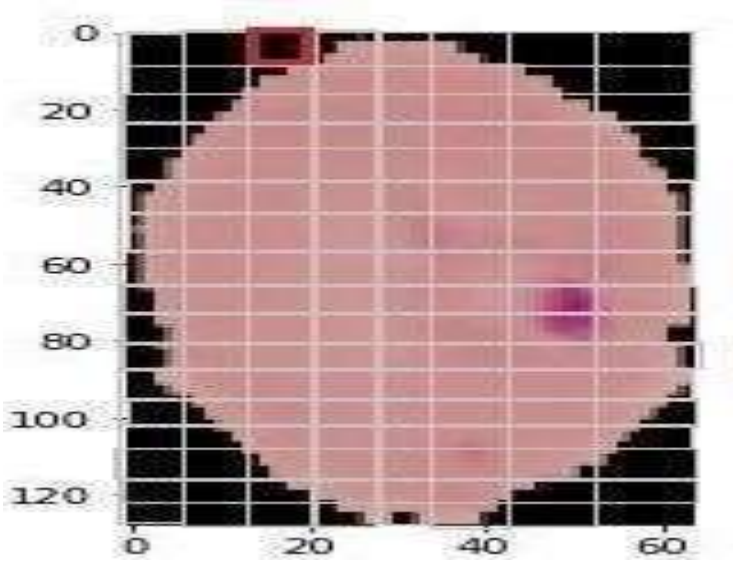

Fig-7. Highlighting single cell next important step is to normalize the histogram. Such that change in brightness won't affect our results.

- Normalize the gradients to (36x1) pixel matrix that is 16×16 cell : As of now, the calculated HOG features for the patch having 8×8 cells of blood smear, the gradients of the image are tactful to the altogether brightness. It is none another than particular portions of the image is brighter compared to other. As it is not possible to eliminate this problem completely but we can reduce it some extent with the help of normalization by considering 2*2 patches having pixel matrix of size 16*16 as shown in fig-8:

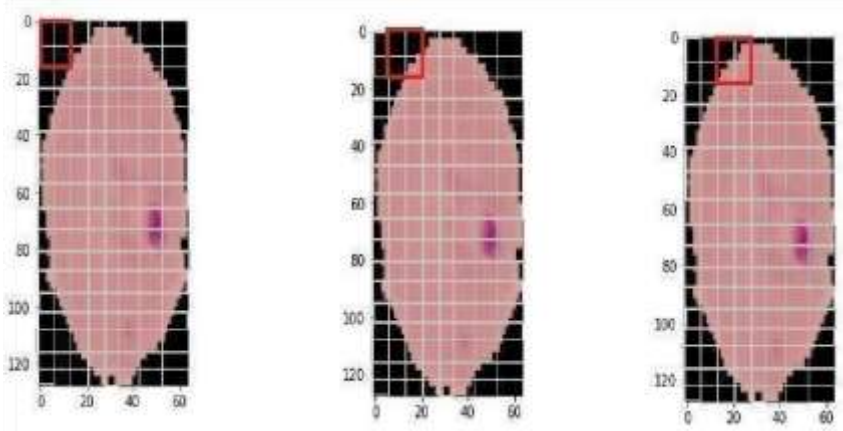

Fig-8 Highlighting (2*2) patches

So here we have to combine four 8x8 matric esto form a big matrix of size 16x16. As of now from one 8x8 matrix we get a vector of size nine and from four matrices we geta vector of size 9*4=36, we can also call this vector as a matrix of36x1. For normalizing we divide the matrix with root of sum of squares of the numbers . Given vec9tormathematically can be written as.

a. V=[b1,b2,b3,….b36]

At first we have to calculate the below equation–(2) where we root value of sum of squares:

b. p= √(b1)2+ (b2)2+ …. +(b35)2+ (b36)2

After getting we use this p to divide each andevery value of vector:

c. Vector after normalization={b1/p,b2/p,b3/p……,b36/p }

The result is a vector which is normalized of 36x1size.

- Features Of Full image: Now we reached the last step for the calculation of Features of Hog. Till now we have generated features of a 16x16 matrix size. Now, we have to combine all of these features to acquire the final image. At first we have to find out total number of such 16×16blockswe would get for a image of 64×128 as shown in fig-9.

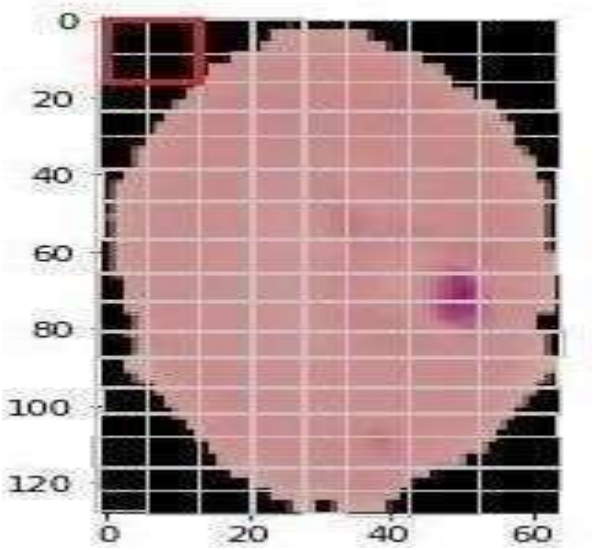

Fig-9. Highlighting (2*2) patches

In horizontal we get seven and in vertical we get fifteen in totalwe get 7x15 =105 pixel matrices of size 16x16 from each matrix we get a vector of size 36x1. So in total we get 36x105=3780 features. From below image we can see the extracted HOG features as shown in fig-10. Finally, we feed Hog_ features for linear SVM for classification of images as Parasitized/Uninfected.

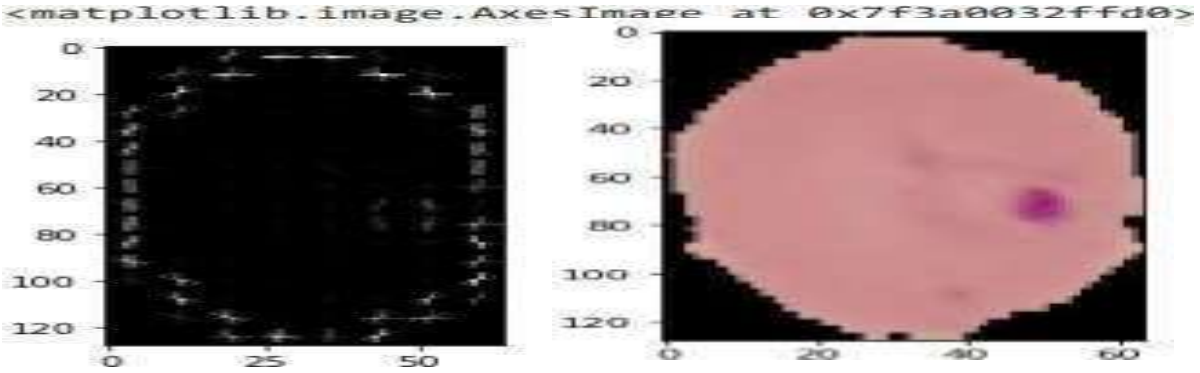

Fig-10. Hog_features and its cell image

4. RESULTS AND CONCLUSION

Other Deep Learning Based model comparatives

FLANN + SSAE MALARIA PARASITE DETECTION METHOD

The trained their model with the FLANN – SSAE techniques for malaria parasite detection from images of blood smears. For detection of malaria parasite, they proposed a CAD model to identify whether the image is taken from a infected person or not [13].

4.1. Experiments and Results

The Support vector Machine (SVM) using HOG features and the proposed CAD model both of them got judged using malaria blood smear images data set.

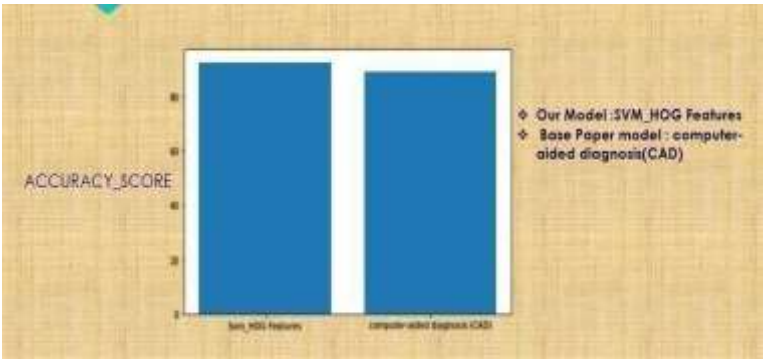

Fig-11 Accuracy-score

TABLE 1. Accuracies deep learning techniques such as Support vector Machine (SVM) using HOG features and proposed CAD scheme.

Deep Learning Techniques	**Accuracy (%)**
Support vector Machine (SVM) using HOG T features	**92.69%**
hProposed CAD e scheme (Base Paper) [13]	**89.10%**

AUC ROC Curve

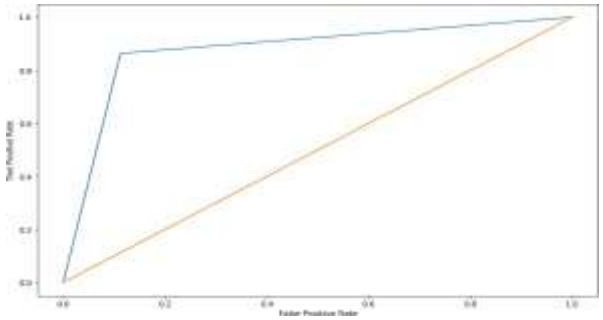

Fig-12. AUC-ROC CURVE

AUC ROC graph or curve the malaria dataset using the Support vector Machine (SVM) using HOG features representing the blood smear image and area under curve give the accuracy of our model as shown in fig- 11.

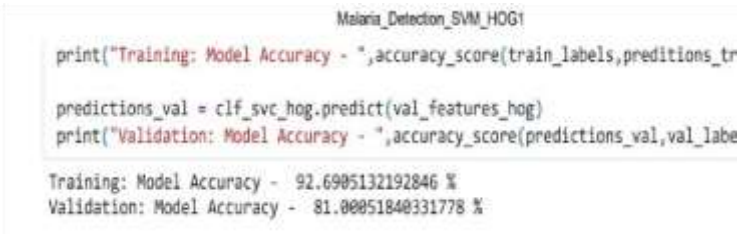

Fig-13. Result obtained using SVM_HOG features

5. CONCLUSION AND FUTURE WORK

The machine learning methods. We used such as HOG Features, SVM have shown limited accuracy for malarial parasite detection from blood smear pictures which is around 92% ,So as part of future work will we try better feature extraction techniques and will train our model on more deep learning techniques and we also try to ensembles different machine learning methods to improve the accuracy.

REFERENCES

[1] World Health Organization. World malaria report 2019 (World Health Organization), 2020.

[2] Poostchi, M., Silamut, K., Maude, R. J., Jaeger, S., & Thoma, G. (2018). Image analysis and machine learning for detecting malaria. *Translational Research*, *194*, 36-55.

[3] Arco, J. E., Górriz, J. M., Ramírez, J., Álvarez, I., & Puntonet, C. G. (2015). Digital image analysis for automatic enumeration of malaria parasites using morphological operations. *Expert Systems with Applications*, *42*(6), 3041-3047.

[4] Makhija, K. S., Maloney, S., & Norton, R. (2015). The utility of serial blood film testing for the diagnosis of malaria. *Pathology*, *47*(1), 68-70.

[5] World Health Organization. (2016). *Malaria microscopy quality assurance manual-version 2*. World Health Organization.

[6] Díaz, G., González, F. A., & Romero, E. (2009). A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. *Journal of biomedical informatics*, *42*(2), 296-307.

[7] Tek, F. B., Dempster, A. G., & Kale, I. (2010). Parasite detection and identification for automated thin blood film malaria diagnosis. *Computer vision and image understanding*, *114*(1), 21-32.

[8] Mustafa, W. A., Santiagoo, R., Jamaluddin, I., Othman, N. S., Khairunizam, W., & Rohani, M. N. K. H. (2018, August). Comparison of detection method on malaria cell images. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)* (pp. 1-6). IEEE.

[9] Umer, M., Sadiq, S., Ahmad, M., Ullah, S., Choi, G. S., & Mehmood, A. (2020). A novel stacked CNN for malarial parasite detection in thin blood smear images. *IEEE Access*, *8*, 93782-93792.

[10] Gopakumar, G. P., Swetha, M., Sai Siva, G., & Sai Subrahmanyam, G. R. K. (2018). Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner. *Journal of biophotonics*, *11*(3), e201700003.

[11] Mehanian, C., Jaiswal, M., Delahunt, C., Thompson, C., Horning, M., Hu, L., ... & Bell, D. (2017). Computer-automated malaria diagnosis and quantitation using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 116-125).

[12] Chowdhury, A. B., Roberson, J., Hukkoo, A., Bodapati, S., & Cappelleri, D. J. (2020). Automated complete blood cell count and malaria pathogen detection using convolution neural network. *IEEE Robotics and Automation Letters*, *5*(2), 1047-1054.

[13] Pattanaik, P. A., Mittal, M., & Khan, M. Z. (2020). Unsupervised deep learning cad scheme for the detection of malaria in blood smear microscopic images. *IEEE Access*, *8*, 94936-94946.

[14] Nguyen, N. D., Bui, D. H., & Tran, X. T. (2019, November). A novel hardware architecture for human detection using HOG-SVM co-optimization. In *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)* (pp. 33-36). IEEE.

[15] Arowolo, M. O., Adebiyi, M. O., Adebiyi, A. A., & Okesola, O. J. (2020). A hybrid heuristic dimensionality reduction methods for classifying malaria vector gene expression data. *IEEE Access*, *8*, 182422-182430.

[16] Wang, S., Yang, C., Preiser, P., & Zheng, Y. (2020). A Photoacoustic-Surface-Acoustic-Wave Sensor for Ring-Stage Malaria Parasite Detection. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *67*(5), 881-885.

Jansson, S., Atkinson, P., Ignell, R., & Brydegaard, M. (2018). First polarimetric investigation of malaria mosquitoes as lidar targets. *IEEE Journal of Selected Topics in Quantum Electronics*, *25*(1), 1-8.

[17] Hole, A. P., & Pulijala, V. (2020). An Inductive-Based Sensitive and Reusable Sensor for the Detection of Malaria. *IEEE Sensors Journal*, *21*(2), 1609-1615.

[18] Kassim, Y. M., Palaniappan, K., Yang, F., Poostchi, M., Palaniappan, N., Maude, R. J., ... & Jaeger, S. (2020). Clustering-based dual deep learning architecture for detecting red blood cells in malaria diagnostic smears. *IEEE Journal of Biomedical and Health Informatics*, *25*(5), 1735-1746.

[19] Chen, K., Perlaki, C., Xiong, A., Preiser, P., & Liu, Q. (2016). Review of surface enhanced Raman spectroscopy for malaria diagnosis and a new approach for the detection of single parasites in the ring stage. *IEEE Journal of Selected Topics in Quantum Electronics*, *22*(4), 179-187.

[20] M. Gupta, H. Wu, S. Arora, A. Gupta, G. Chaudhary, and Q. Hua, "Gene Mutation Classification Through Text Evidences Facilitating Cancer Tumour Detection," *Journal of Healthcare Engineering, Hindawi, pp. 1-16, vol. 2021, Doi:* https://doi.org/10.1155/2021/8689873.

[21] R. Jain, M. Gupta, S. Taneja, and J. Hemanth, "Deep Learning based detection and analysis of COVID-19 on Chest X-Ray Images," Applied Intelligence (APIN), 10.1007/s10489-020-01902-1, APIN-D-20-01185R3, Oct 2020