cisdi Journal

Computing, Information Systems & Development Informatics Journal
Vol. 11 No. 1, 2020 - www.isteams.net/cisdijournal

Image Processing and Supervised Learning for Efficient Detection of Animal Diseases

[1]Akpojaro, J. & [2]Bello, R-W.
[1, 2]Department of Mathematical Sciences, University of Africa, Toru-Orua, Bayelsa State, Nigeria
[2]School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau, Pinang, Malaysia
E-mails: [1]jackson.akpojaro@uat.edu.ng, [2]sirbrw@yahoo.com
Phones: [1](+2348138803299), [2](+2348038897840, +601136671473)

ABSTRACT

Animal disease management such as trypanosomiasis can be a frustrating proposition if proper techniques are not employed for the disease management as lack of information on diseases that can affect individual animals or an entire herd is tantamount to economic loss. Classifiable symptoms associated with the disease help strategize for problem identification and solution provision. Currently, disease management, symptoms classification and diagnoses are manually performed and are time consuming due to the fact that it takes longer time for infected animal to be manually diagnosed especially those animals which are remote from veterinary. These challenges motivate the development of detection tools that can perform automatically using deep learning approaches such as convolutional neural networks which have received great acceptance in literature but are computationally expensive and consumes huge amount of training data. This paper seeks to improve on the deep learning methods of managing animal diseases by using non-deep methods such as random forests and support vector machines with adoption of pre-processing and thorough testing methodology. Test carried out on 1000 images shows random forests system performing better with achievement of 95.20% accuracy. The system's accuracy, recall, and precision are higher than the previous non-deep approaches and performing convolutional neural networks approach. To our best knowledge, this work is one of the newest works carried out to facilitate the detection of animal diseases for the benefit of animal husbandry; this implies that more research efforts are ongoing to improve this area of research for reliable policies and practices. Unavailability of huge dataset is the greatest challenge of the research which limited the extent to which comparison of the proposed method to others could have been made possible.

Keywords: Animal, Convolutional neural networks, Disease, Management, Support vector machines, Trypanosomiasis, Random forests

CISDI Journal Reference Format
Akpojaro, J & Bello, R-W. (2020): Image Processing and Supervised Learning for Efficient Detection of Animal Diseases. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 11 No 1, Pp 39-48.
Available online at www.computing-infosystemsjournal.info

1. INTRODUCTION

Trypanosomiasis, otherwise known as nagana, or sleeping sickness is a vertebrate disease that can have annihilating effects, not only on individual animals who contract it, but also on the owners of the animals in term of economic loss due to low in productivity and even death. Trypanosomes are the protozoan parasites which are transmitted by biting insects (tsetse flies) through their saliva and infect the blood of vertebrates and they are of several species in the taxonomy transmitted by the biting insects (WHO, 2017). Batista *et al.,* (2011), nagana is caused by Trypanosoma mainly in the continent of Africa which is now traceable to South America. The blood of host

is infected by the trypanosomes causing weakness, fever and lassitude, which can lead to a state of comatose torpor, weakness characterized by a lack of vitality or energy. Different breeds of cattle from West Africa such as N'Dama's Bos taurus breed show noteworthy tolerance to nagana pathology in contrast with the breeds of cattle from east Africa such as Zebu's Bos indicus breed that are susceptible to the nagana pathology (Courtin *et al.,* 2008).

The transmission of the disease is by trypanosomes which are developed in tsetse flies, about one to a few weeks in its biological vector. When an infected tsetse fly (Figure 1) bites an animal, the parasites are transmitted. Significant weight loss and anemia are some of the likely signs and symptoms demonstrate by the infected vertebrates (Figure 2). Various other symptoms are observed; these include fever, adenitis, oedema, nervous disorders and dermatitis. To eliminate trypanosomes completely, the immune system of animal must be able to secure against the disease but the immune system of animal due to one reason or another might be unable to eliminate trypanosomes completely making the animal to become unobvious carriers. If the low in immune system animal is stressed, these unobvious infections can be reactivated (Odeniran and Ademola, 2018).

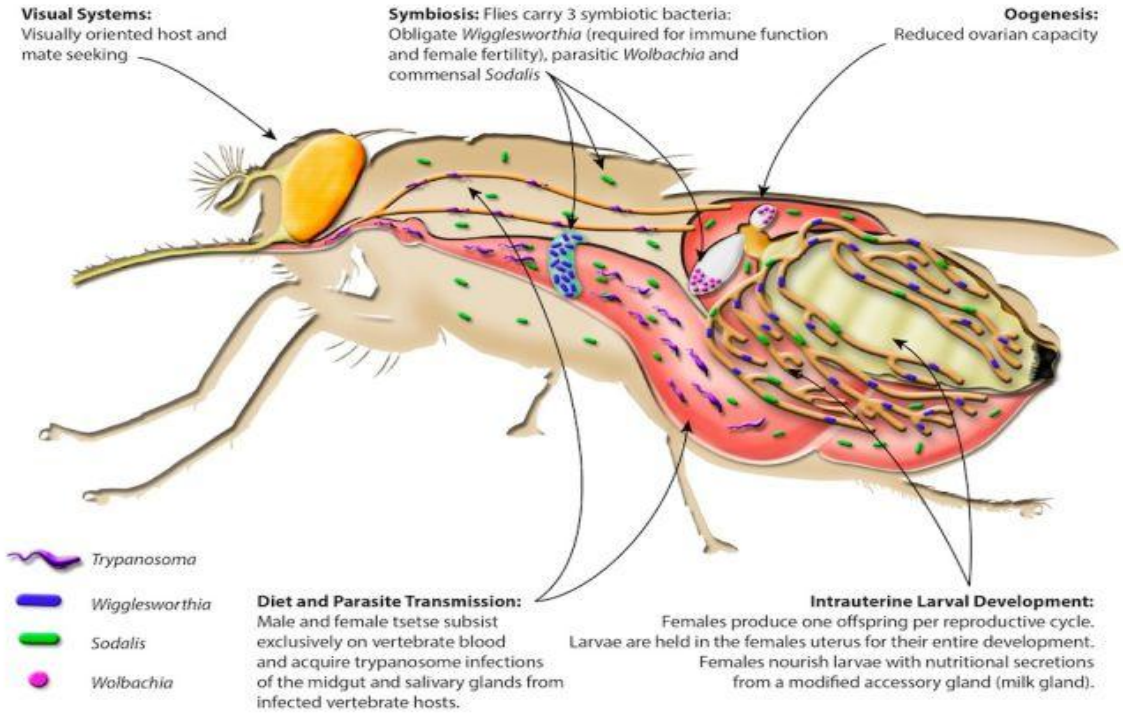

Figure 1: Portrait of the Blood-Sucking Tsetse Fly (livescience.com)

According to Finelle (1974) and Karshima *et al.,* (2016), the disease cannot be diagnosed with certainty except by blood microscopic examination or various serological reactions. This presents a clear problem of the infected animals not having timely veterinary attention as the available microscopy standard for diagnosing trypanosomes is performed by expert who are either far from the locations where the animals are reared and sometimes, the procedure is time consuming and costly (Shillcutt *et al.,* 2008). These challenges motivate the need for cheap and reliable automated trypanosomes diagnosis and detection systems that mitigate the time burden on veterinary experts.

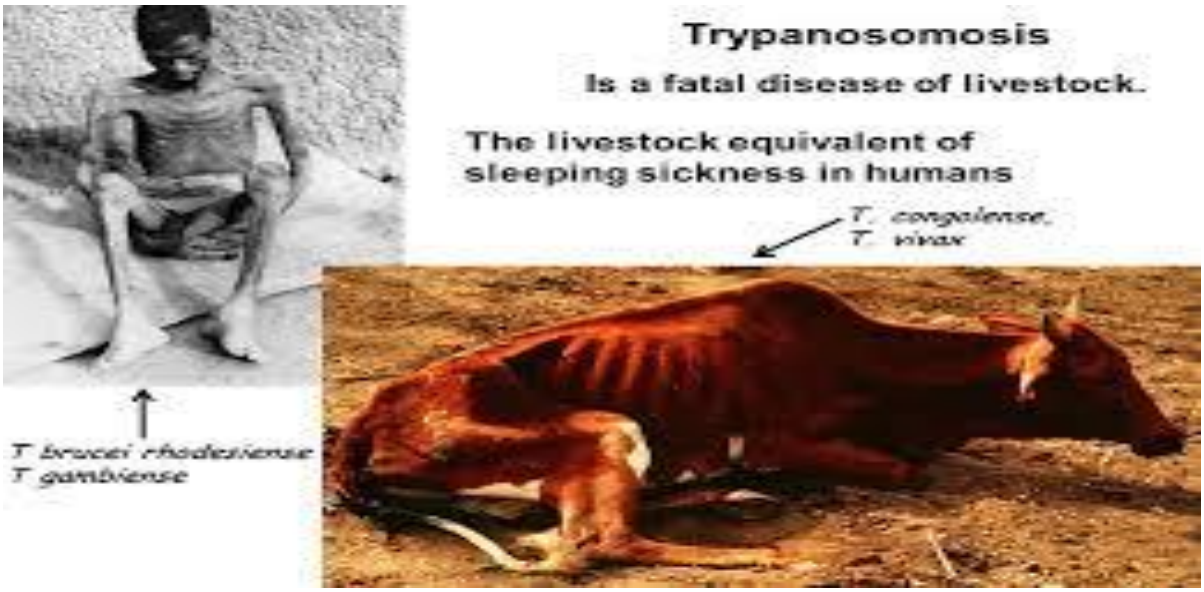

Figure 2: Trypanosomiasis Infected Vertebrates (slideplayer.com)

Existing literatures have not exhaustively investigated the limitations of non-deep machine learning approaches in detecting trypanosomes making the approaches proved to be more suitable for use in regions with limited computing resources. Presented in this paper are two non-deep supervised learning systems namely, random forests and support vector machines. The adoption of strict testing on a fairly large dataset makes this paper differs from any existing work using non-deep models. The rest of the paper is as follows: Materials and methods are presented in section 2 with sub-sections; results and discussion are presented in section 3; the conclusion and future work are presented in section 4.

2. MATERIALS AND METHODS

Discussed here are the morphology, software framework, algorithms for feature extraction, dataset used, criteria for evaluation, computational equipment, the experimental design, support vector machine algorithm, and random forests.

2.1 Morphology

Knowing that the identification of individual species of various trypanosomes and their exact cause depends on sound knowledge and understanding of the basic features of the protozoan parasites, we carried out the theoretical study of the basic features of the protozoan parasites. This enables us to appreciate the basic features possessed by all trypanosomes, which in-turn enables the recognition of diagnostic differences and the identification of species.
The diagrammatical description of the fundamental features of trypanosome is shown in Figure 3. The parasite comprises of a single cell of various sizes from 8 µm to more than 50 µm.

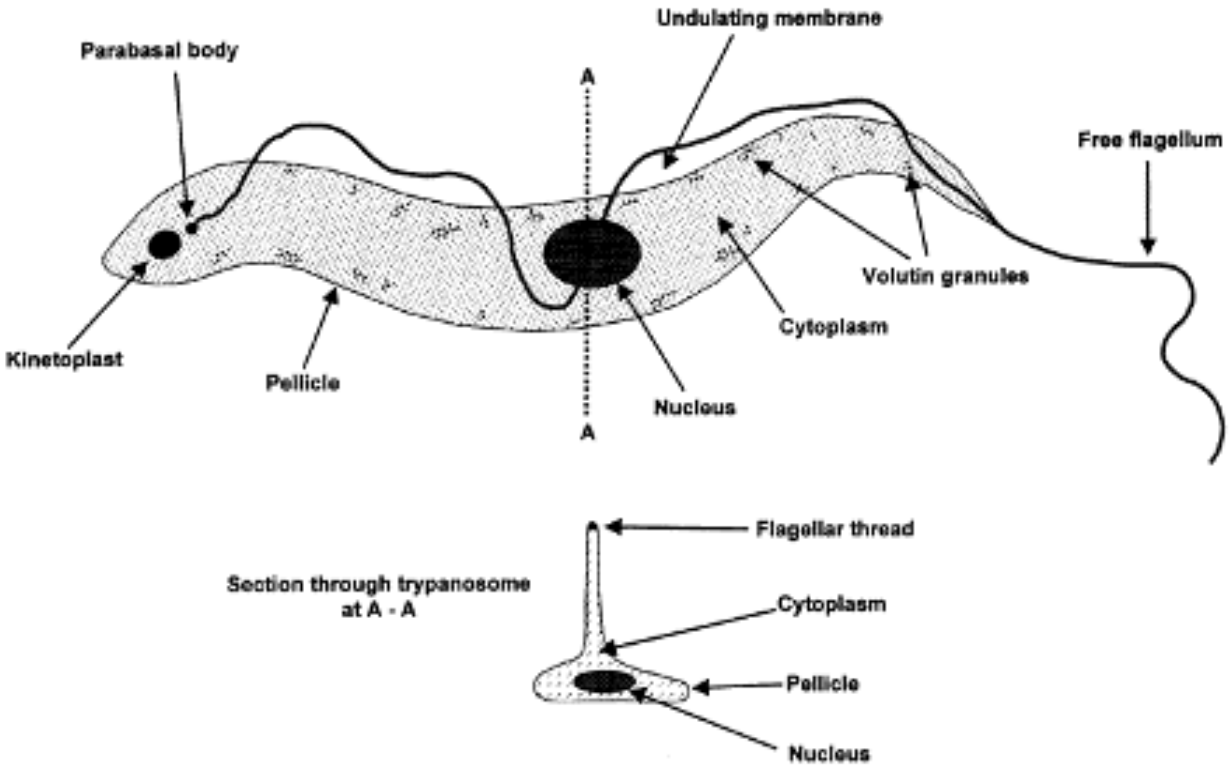

Figure 3: Diagram of Trypanosome (Uilenberg, 1998)

2.2 Software Framework

MATLAB R2016b toolbox running vision and image processing is employed for the work optimisation. To make a complete system, personal computer running operating system Microsoft Windows 10 and Microsoft Visual Studio 2016 is employed. MATLAB R2016b provides image processing and computer vision Toolbox™ through which a comprehensive set of reference-standard algorithms and workflow apps for vision and image processing is provided. This work optimises MATLAB R2016b in the image segmentation, and object identification using the techniques of deep learning. Through computer vision apps, ground truth labelling and camera calibration workflows are automated. C/C++ code generation is supported by many MATLAB R2016b toolbox functions for desktop prototyping and embedded vision system deployment.

2.3 Algorithms for Feature Extraction

In order to reduce the input dimensionality, feature extraction algorithms were applied prior to when input was supplied to machine models making the models to run much faster and perform better. The three approaches considered for this work are:

- Hu moments as proposed by (Hu, 1962). The approach is a set of seven moment invariants, which are all invariant to rotation, scale, and translation.
- Haralick texture attributes as proposed by (Haralick *et al.* 1973). The approach is a set of 14 textural features that can be extracted from images in order to improve image classification accuracy.
- Histograms, through which the number of pixels that fall into a specified number of intensity bins, are counted, for each colour channel.

2.4 Datasets

1000 datasets were used to run the experiments. The datasets were made up of blood cell images that have been pre-cropped, with 500 placed under the infected classification and another 500 images placed under the uninfected classification, resulting to 1000 images.

2.5 Criteria for Evaluation

Three criteria were used to evaluate the systems, they are: (1) accuracy (2) recall (3) precision. When correct prediction is made by a system, such overall ability displayed by the system is referred to as accuracy. The avoidance of false negative by the system is referred to as recall. The avoidance of false positive by the system is referred to as precision. They are mathematically represented as follows:

$$Accuracy = \frac{(AnB)}{(AuB)} \times 100 \quad (1)$$

Where A is the blood image that is segmented manually and B is the image that is cropped and segmented dynamically.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

2.6 Computational Equipment

The development, the training and the testing were all carried out on the same machine: the detection system runs on the following system specifications: Intel® Core™ i5 processor, 8GB of RAM, dedicated graphics card, 1 terabyte of hard disk space.

2.7 Experimental Design

System performance evaluation was based on the ability of the system to operate on data belonging to the same dataset used for training. Five iterations produced the results got from this experiment with each time trained on a batch size of 100 images and tested on a disjoint set of 1000 images. Indicated by the initial testing was the fact that a training set of 100 images was enough for the convergence of the models, so a more extensive testing set could be formed using the rest of the data. The essence of this experiment is to evaluate the enablement of the system to perfectly predict infection in images collected in the same manner to the training data.

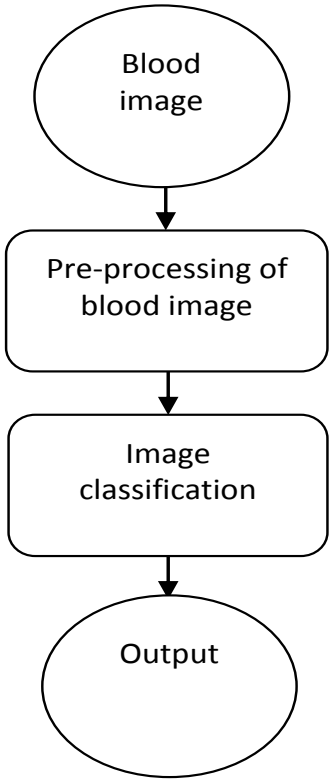

Figure 4: Proposed Method

As shown in Figure 4, the system accepts the blood image (Figure 5), performs pre-processing operations such as colour and texture features extraction, cropping and segmentation on the image to remove unwanted objects such as background patches and other heterogeneous objects, and to improve the image quality for classification purpose by the classifiers.

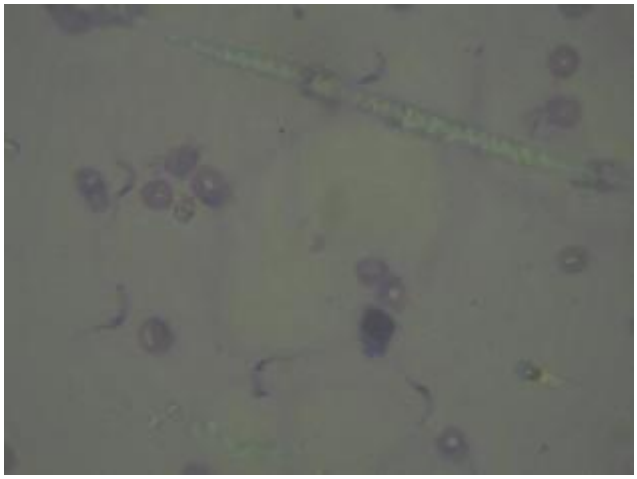

Figure 5: Image of Trypanosomes in a Giemsa Stained Blood Smear

2.8 Support Vector Machines (SVM) Algorithm

After the Histogram of Gradient feature vectors have been extracted, an algorithm for classification would be applied for the detection of the target parasite. Given a set of training database with the size N, we express it as follow:

$$(x_i, y_i) 1 \leq I \leq N \quad (4)$$

Where,
$x_i \in R^n$ is a kind of descriptor vector that belongs to the class labelled by $y_i \in \{-1, 1\}$, the class indicates the binary result of the training database, −1 stands for false or negative, 1 stands for true or positive. The aim of the classification is to get constructed the equation of a hyper-plane which separates the database set in such a way that all the points having the same value of yi = 1 and yi = −1 are grouped together on either sides of the hyper plane.
In other words, w and b that satisfy the following inequality need to be found:

$$y_i (w.x_i + b) > 0, i = 1, ..., N \quad (5)$$

It is noted that there is neither uniqueness nor existence of the hyper planes. We can conclude that the sample set is linearly separable when there is existence of a hyper plane that satisfies equation (5) making the update of w and b always possible in such a case such that:

$$1 \leq i \leq N \; y_i (w.x_i + b) \geq 1, i = 1, .., N \quad (6)$$

Based on equation (6), 1/ ||w|| is the distance between the closest points to the hyper plane. So, equation (5) can be changed to:

$$y_i (w.x_i + b) \geq 1 \; \forall i \quad (7)$$

2.9 Random Forests

Random forest is an ensemble tree-based learning algorithm (Figure 6). By ensemble algorithm, we mean it is an algorithm which combines more than one algorithms of same kind or different kind for the classification of objects. This happens when running prediction over support vector machines and decision tree and then taking vote for final consideration of class for test object. The random forest classifier is a set of decision trees from randomly selected subset of training set. It aggregates the votes from different decision trees to decide the final class of the test object.

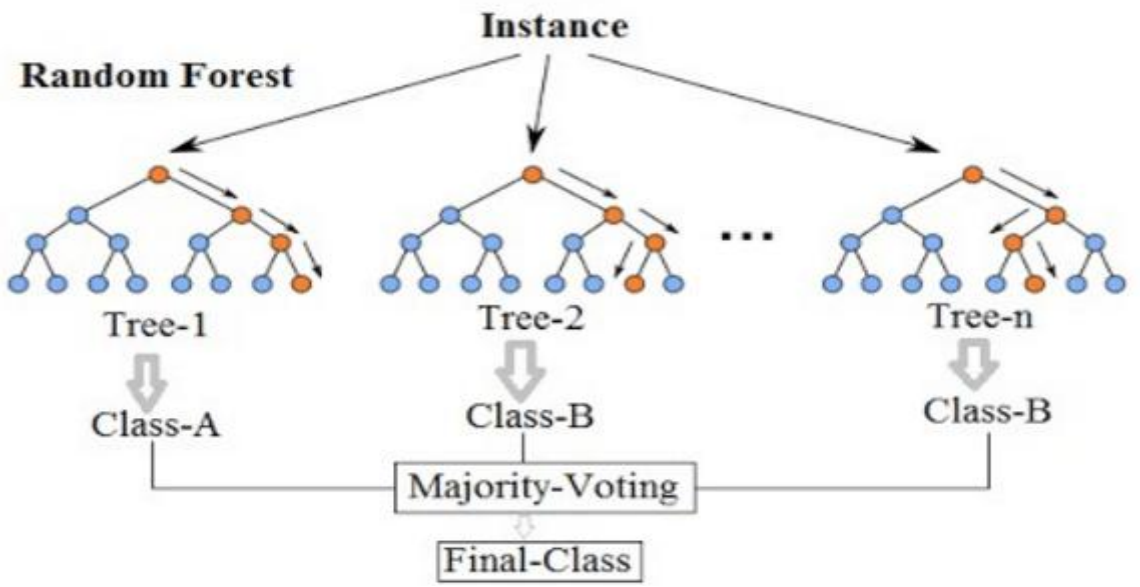

Figure 6: Structure of Random Forest Classification (Chakure, 2019)

3. RESULTS AND DISCUSSIONS

Accuracy of 92.30%, recall of 94.06% and precision of 90.28% were achieved by support vector machines (Table 1). Accuracy of 95.20%, recall of 95.15% and precision of 95.10% were achieved by the random forests (Table 1). Haralick *et al.,* (1973), to attain higher recall while comparable overall accuracy is maintained is better than to attain higher precision in the context of initial screening tests. This is due to the fact that it is far more necessary to guide against the occurrence of false negative, as this could lead to infected animals not being attended to for treatment when the need arises. Another test with higher precision may be performed to ensure overall accuracy by following the high sensitivity screening test. An automated trypanosomes screening system could identify cells with a high probability of infection and present these to a pathologist for confirmation. This minimizes the number of cells which must be checked by the expert while still ensuring diagnostic precision.

Deep learning as a new trend in machine learning and extension of multi-layer neural network classifiers with back-propagation ability for training has boosted many non-medical areas in term of performance. Two different kinds of layers are used in successions which handle the large training sets. Large annotated training datasets are required in deep learning; this has made research in medical applications using deep learning a difficult task because of knowledge expertise and privacy associated with medical related datasets. Liang *et al*., (2016), are the first people to apply deep learning to medical diagnosis issue by using conventional cell segmentation approach and convolutional neural network to differentiate between infected and uninfected cells in thin blood smears.

This technological breakthrough in medical applications is followed by Dong *et al.,* (2017), Dong *et al.,* (2017) and Gopakumar *et al.,* (2018), used convolutional neural networks, Bibin *et al*., (2017), used deep belief networks, and recently Hung et al., (2017), presented an end-to-end framework using faster Region-based Convolutional Neural Network. Handcrafted features are not required in deep learning making it one of the best feature learning and representation techniques.

Table 1: Experimental Results (%)

System	Performance Evaluation		
	Accuracy	***Recall***	***Precision***
Random Forests	95.20	95.15	95.10
Support Vector Machines	92.30	94.06	90.28

4. CONCLUSION AND FUTURE WORK

Presented in this paper are experiments that demonstrated non-deep supervised learning techniques as alternative techniques to widely known deep learning approaches for trypanosomes classification. Moreover, automated detection of trypanosomes using the random forests and support vector machines has shown improvement over the existing methods in terms of accuracy, recall, and precision. Test carried out on 1000 images shows random forests system performing better with achievement of 95.20% accuracy. The system's accuracy, recall, and precision are higher than the previous non-deep approaches and performing convolutional neural networks approach.

To our best knowledge, this work is one of the newest works carried out to facilitate the detection of animal diseases for the benefit of animal husbandry; this implies that more research efforts are ongoing to improve this area of research for reliable policies and practices. Unavailability of huge dataset is the greatest challenge of the research which limited the extent to which comparison of the proposed method to others could have been made possible. Timely detection and diagnosis of the disease using the proposed method can help in ensuring early treatment of the unobvious infected animals and devising a way of eradicating the disease. This work is not completely excluded from the notable disadvantage attributed to random forests in accommodating overfitting for some datasets with noisy classification/regression tasks.

REFERENCES

1. Batista, J. S., Rodrigues, C. M., García, H. A., Bezerra, F. S., Olinda, R. G., Teixeira, M. M., and Soto-Blanco, B. (2011). Association of Trypanosoma vivax in extracellular sites with central nervous system lesions and changes in cerebrospinal fluid in experimentally infected goats. *Veterinary research*, 42:63; 1-7.
2. Bibin, D., Nair, M. S., and Punitha, P. (2017). Malaria parasite detection from peripheral blood smear images using deep belief networks. *IEEE Access*, 5: 9099-9108.
3. Chakure, A. (2019). Random Forest Classification and its implementation in Python. Towards Data Sciences. Retrived 4th April, 2020 from https://towardsdatascience.com/random-forest-classification-and-its-implementation-d5d840dbead0.
4. Courtin, D., Berthier, D., Thevenon, S., Dayo, G. K., Garcia, A., and Bucheton, B. (2008). Host genetics in African trypanosomiasis. *Infection, Genetics and Evolution,* 8(3), 229-238.
5. Dong, Y., Jiang, Z., Shen, H., Pan, W. D., Williams, L. A., Reddy, V. V., ..., and Bryan, A. W. (2017). Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 101-104.
6. Dong, Y., Jiang, Z., Shen, H., and Pan, W. D. (2017). Classification accuracies of malaria infected cells using deep convolutional neural networks based on decompressed images. In Southeast Con 2017, IEEE, pp. 1-6. .

7. Finelle, P. (1974). African animal trypanosomiasis. IV. Economic problems. World Animal Review. Retrieved 4th March, 2020 from http://agris.fao.org/agris-search/
8. Gopakumar, G. P., Swetha, M., Sai Siva, G., and 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.
9. Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), 610-621.
10. Hu, M. K. (1962). Visual pattern recognition by moment invariants. IRE transactions on information theory, 8(2), 179-187.
11. Hung, J., and Carpenter, A. (2017). Applying faster R-CNN for object detection on malaria images. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 56-61).
12. Karshima, S. N., Lawal, I. A., and Okubanjo, O. O. (2016). Feeding patterns and xenomonitoring of trypanosomes among tsetse flies around the Gashaka-Gumti National Park in Nigeria. *Journal of Parasitology Research,* Volume 2016: 1 – 7.
13. Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., ..., and Huang, J. X. (2016). CNN-based image analysis for malaria diagnosis. In 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 493-496.
14. Odeniran, P. O., and Ademola, I. O. (2018). A Meta-analysis of the Prevalence of African Animal Trypanosomiasis in Nigeria from 1960 to 2017. *Parasites & Vectors*, 11:280; 1-12.
15. Shillcutt, S., Morel, C., Goodman, C., Coleman, P., Bell, D., Whitty, C. J., and Mills, A. (2008). Cost-effectiveness of malaria diagnostic methods in sub-Saharan Africa in an era of combination therapy. Bulletin of the World Health Organization, 86: 101-110.
16. Uilenberg, G. (1998). A field guide for the Diagnosis, Treatment and Prevention of African Animal Trypanosomosis. Food and Agriculture Organization of The United Nations, Rome, FAO 1998.
17. WHO (2017). Report of the second WHO stakeholders meeting on rhodesiense human African trypanosomiasis (sleeping sickness). Retrieved 01 February, 2020 from https://apps.who.int/iris/bitstream/handle/