ISSN (E): 2959-3921 *Ethiopian International Journal of Engineering and Technology (EIJET)*
(P): 2959-393X *Volume 1, Issue 2, 2023*
DOI: https://doi.org/10.59122/144CFC14

Cattle Pinkeye Disease Classification Using Machine Learning

Serkalem Mekonnen[1] and Mohammed Abebe[2*]

[1]Department of Computer Science, Wolaita Sodo University, Ethiopia

[2]Faculty of Computing & Software Engineering, Arba Minch University, Ethiopia

[*]Corresponding Author Email: mohammed.abebe@amu.edu.et

Abstract

Pinkeye (infectious bovine keratoconjunctivitis, or IBK) is a bacterial infection of the cattle eye that causes inflammation and, in severe cases, temporary or permanent blindness. It is a painful, debilitating condition that can severely affect animal productivity. Due to lower weight gain, lower milk production, and higher medical costs, the cattle industry could experience large losses. Previous studies tried to classify livestock diseases using machine learning, but there has been a lack of studies conducted on pinkeye disease classification. The proposed study aims to design a classification model to classify whether the infected cattle have pinkeye or not at an early stage by analyzing a set of attributes. The study collected data from the Wolaita Sodo Kenido Koyisha Wereda Livestock and Fishery Office. The significance of this study is to prevent the expansion of disease among the cattle with early detection for taking precautionary measures. The researchers used the percentage splits 80/20, 70/30, 60/40, and 90/10 to build classification models. Based on the results of the experiments, the researchers chose the 70/30 split due to the better performance obtained. The study trained four different models, including Random Forest, AdaBoost, Artificial Neural Network, and Extreme Gradient Boost algorithms. These models were selected based on an exhaustive study conducted. To assess the algorithm's performance, confusion matrix, accuracy, precision, recall, and f1-score have been utilized. With a 99.15% accuracy, the Artificial Neural Network outperforms the other algorithms by all the metrics except recall.

Keywords: Cattle Pinkeye, IBK, Infectious Bovine Keratoconjunctivitis, Machine Learning, Pinkeye Disease Classification.

I. Introduction

Ethiopia is one of the nations in Africa with the highest number of livestock (over 65 million) [1]. Pinkeye also known as infectious bovine keratoconjunctivitis (IBK), is a frequent cattle disease that produces redness and ulceration in the eye. Pinkeye is a serious eye condition that has an economic influence on animals' performance [2]. It is exceedingly painful and causes huge economic losses due to decreased

***Received:** July 26, 2023; **Revised**: 17 September 2023; **Accepted:** 04 October 2023; **Published:** 31 December 2023.*
*Corresponding author- **Dr. Mohammed Abebe***

weaning weights and livestock value. Pinkeye is caused mostly by the bacteria Moraxella bovis, but it can also be caused by other bacteria and viruses. Environmental variables that increase the risk of pinkeye in cattle include seed heads, dust, pollen, and UV light. These irritants scrape the cornea of the eye allowing the bacteria to attach more easily. These irritants promote an increase in tear production, which attracts face flies, which can spread the bacteria that causes pinkeye [3]. Pinkeye can damage up to 80% of a herd, resulting in weaner calves losing 10% of their total weight. Cattle may die from malnutrition, thirst, or accidents if both eyes are affected with pinkeye. Damage to the eye can sometimes be serious enough to cause lifelong blindness [4]. The degree of inflammation may eventually become severe enough to produce corneal ulceration, which may rupture, potentially ending in blindness.

While not lethal, pinkeye has a significant financial impact on the Ethiopian cattle sector. Pinkeye can have an impact on the prices received for cattle at sale due to price reductions, in addition to the fact that calves weigh 36–40 pounds less at weaning [5]. Furthermore, pinkeye is a costly condition for beef producers. Poor weight gain and appetite reduction in animals affected by visual impairment and ocular pain result in enormous losses [6]. Farms may benefit from identifying cows who are more likely to develop health issues such as clinical mastitis, subclinical ketosis, lameness, and metritis to promptly avoid and treat their harmful impacts [7]. However, allowing serious cases to continue to a severe stage without treatment is poor management and unacceptable from a welfare standpoint. Therefore, this study proposed the development of a machine-learning model for the early detection of cattle pinkeye disease. The proposed model can be able to predict whether the cattle have pinkeye or not at an early stage by analyzing selected attributes from the clinical dataset using machine learning so that appropriate medical decisions can be made to avoid severe damage.

II. Literature Review

Machine learning (ML) is a type of artificial intelligence (AI) that uses historical data as input to forecast new output values. It is a discipline of computer science concerned with the study and interpretation of data patterns and structures to enable learning, reasoning, and decision-making without the involvement of humans. Machine learning allows a user to submit large volumes of data to a computer algorithm, which analyzes the data and produces data-driven recommendations and decisions based only on the data provided. ML plays a critical function in accurately extracting information from the massive amount of data at that point. Various studies have been conducted on the development of assistive tools for cattle diseases. However, only little effort is undertaken in the Ethiopian context. There are various challenges endured by the cattle industry in Ethiopia due to the acute shortage of assistive tools especially in rural areas. The

Received: July 26, 2023; **Revised**: 17 September 2023; **Accepted:** 04 October 2023; **Published:** 31 December 2023.
Corresponding author- **Dr. Mohammed Abebe**

challenges include economic loss due to lower weight gain, decreased milk supply, and so on. It is because of these facts that the researchers believe that AI-enabled tools play a paramount role in alleviating some of the aforementioned challenges. Some of the related works are discussed hereunder.

The study "Application of Artificial Intelligence for Livestock Disease Prediction" used artificial intelligence and Geographic Information System (GIS) to create disease-climate association models to anticipate 13 economically critical livestock disease outbreaks in India. Data on disease outbreaks was gathered from 31 AICRP (All India coordinated research project) centers. The study used two regression models, Generalized Linear Models (GLM) and Generalized Additive Models (GAM). They used six machine-learning algorithms to predict livestock diseases [8]. The study of D. Ashar [9], presented a system to detect the existence of livestock diseases to take precautionary measures and notify the livestock owner if the condition is likely to cause a sudden death. The goal was to raise awareness of an illness that can bring death unexpectedly at any moment in the future. The study developed a machine learning model using support vector machines (SVM). The model classifies diseases based on the information entered by the user. The user can also choose whether to see the website in Hindi or English. The research "Mobile-based Cattle Infectious Disease Prediction System" [10] proposes a novel prediction method for identifying infectious illness in cattle with the help of a mobile-based information system. This research primarily used Naïve Bayes classifiers to categorize the level of risk presented to cattle by evaluating six baseline animal health syndrome patterns.

Another study [11] was conducted to identify the on-farm risk factors associated with pinkeye disease in Australian cattle. Data were gathered from cattle farmers using a custom-designed online questionnaire. Results revealed that farm location, farm grazing area, farmer-reported dust levels, fly levels, rain levels, animal zebu content, and cattle age were significantly associated with pinkeye prevalence. The results confirm that pinkeye disease is multifactorial and is associated with a range of host and environmental factors. The study suggested these findings should be used to assist in the control of the disease and improve pinkeye outcomes in Australian cattle. According to the model and sample preparation technique utilized, a study produced two biomarker models that correctly categorized the M. bovis bacteria according to genotype with an overall accuracy ranging from 85.8 to 100%. These models offer a useful resource for investigations of genetic connections with disease, allowing epidemiological research at the level of subspecies, and can be applied to improve disease prevention methods. In this study, strain classification was the sole use of genetic indicators [12]. Finally, to the best of the researchers' knowledge, an exhaustive

***Received:** July 26, 2023; **Revised**: 17 September 2023; **Accepted:** 04 October 2023; **Published:** 31 December 2023.*
*Corresponding author- **Dr. Mohammed Abebe***

ISSN (E): 2959-3921 Ethiopian International Journal of Engineering and Technology (EIJET)
(P): 2959-393X Volume 1, Issue 2, 2023
DOI: https://doi.org/10.59122/144CFC14

literature survey showed that there are no machine-learning models developed for the early detection of pinkeye, hence the originality and significance of the study is justified.

III. Materials and Methods

A. Research Process

Finding a research problem is the first step in any research process. The context of the research problem is then discovered through a literature review. Moreover, to express the research questions and to obtain a deeper comprehension of the subject, a literature survey was conducted to find out the justifiable research gap. Based on the identified research gaps, the problem was formulated. After problem formulation, the research questions, objectives, and hypotheses were articulated. Solving the issue entails putting out and creating a plan for data collection, analysis, and experimentation. The formation of a research study design involves choosing a sample size and gathering data from it. The suggested solutions are put into practice and assessed after the processing and analysis of the gathered data and the determined classification model. The researchers used Open-Source tools for data analysis, model design, and validation. Specifically, the study used Python machine-learning libraries for overall model development and testing. The overall research process followed to undertake this research study is shown in Fig. 1. The goal of this research was to build a pinkeye disease classification model based on data collected from clinical trials. Models were selected based on a detailed experimental comparison, their popularity in various machine learning classification tasks, and their performance in previous research. Artificial Neural Networks (ANN), Random Forest, XGBoost, and Adaboost were used in this work for comparative analysis of supervised machine learning classification for pinkeye disease classification.

B. Data Source and Data Collection

Data collection and preparation techniques aid in the generation of high-quality data and the enhancement of classification outputs. To train the proposed model for accurate classification of cattle pinkeye disease, a large number of eye illness records with complete information is required. In this study, we collected a dataset from Wolaita Sodo Kenido Koysha Wereda Livestock and Fishery office, which contains 5508 eye illness records with 22 distinct features collected between the years 2007 to 2022. The researchers transformed the original data into digital formats to make it suitable for machine learning. The data in this study was gathered in a variety of ways. Primary and secondary data sources were utilized. The main source of the data was secondary data, which was collected from clinical records, the researchers also employed primary sources (interviews) for a better understanding of the research domain. The total size of the dataset

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

after preprocessing was 5508 instances with 21 independent attributes and a target class (has pinkeye or not). Out of the total dataset, 3377 records were cattle affected with pinkeye disease and the remaining 2131 are cattle that were not affected by the disease. The dataset had a class imbalance. However, instead of adding synthetic data or removing data, the study opted to use various performance measurement techniques to make sure that the model is not biased towards the class with more observations.

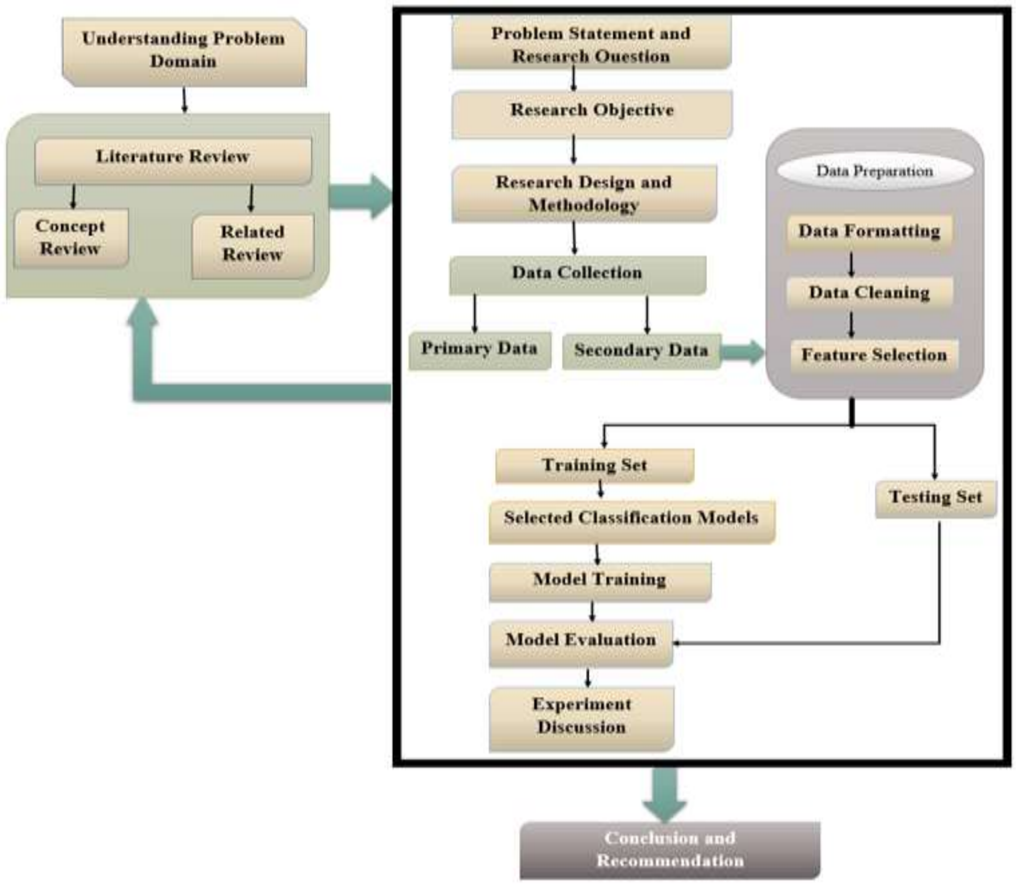

Fig. 1. Overall Research Process

C. Feature Selection

A strategy for selecting, identifying, and eliminating attributes from a dataset to make it suitable for machine learning algorithms is called attribute selection (feature selection). It is a task of selecting a minimal number of features/attributes that are sufficient for correctly classifying the target labels. Feature selection increases the efficacy of the classification model while reducing the computational complexity of the model. Selecting model input variables is an important step in creating an accurate model in machine learning. The study applied feature importance analysis for feature selection.

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

The feature importance analysis refers to methods for scoring each input feature for a certain model. The scores simply indicate the importance of each feature to the target variable. There are various features and important analysis techniques. Out of those methods, the permutation feature importance method was used in this study. Using the permutation importance technique, we run the model while randomly shuffling the values of a single column to evaluate how the scores change. If the scores were significantly impacted, the feature was found very crucial to the model; otherwise, it does not significantly improve the model's performance. A higher score of 0.7 in the dataset indicated that the particular feature has more impact on the model. The importance scores of other features in the dataset were compared to determine the rank of each feature. Fig. 2 presents feature importance scores of the independent features in descending order. Fig. 2 shows the results of fitting a Random Forest model and accumulating the computed permutation feature significance scores.

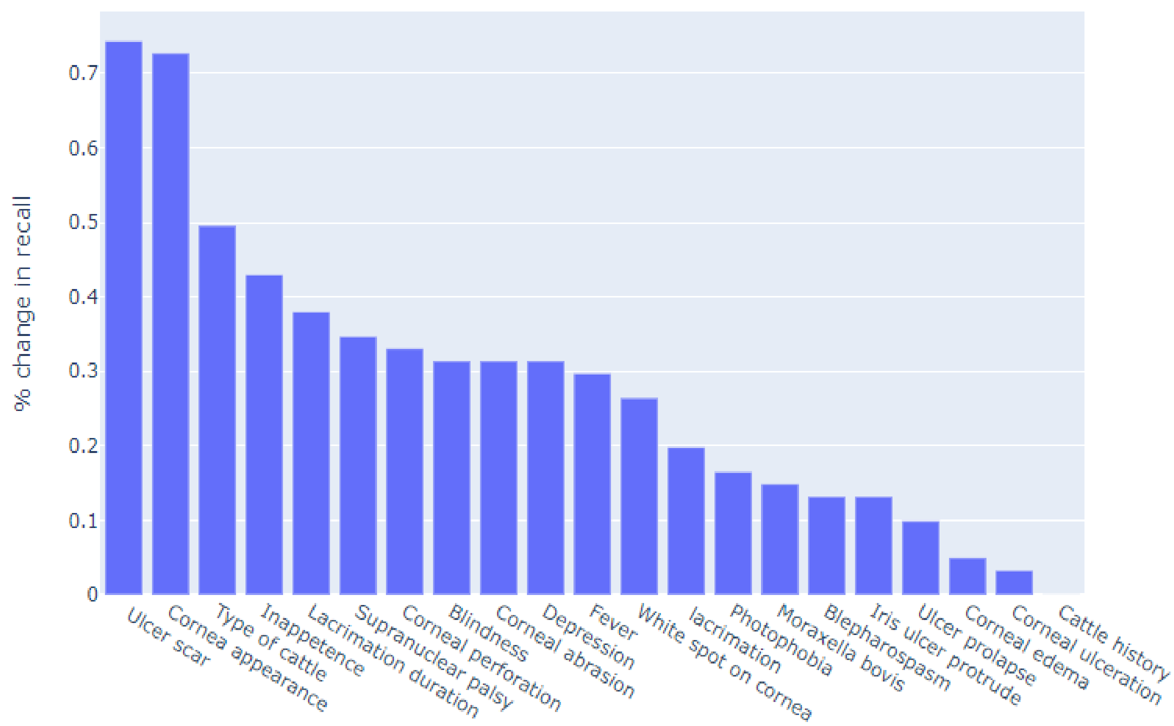

Fig. 2. Permutation Feature Importance Analysis Results

Each feature in the dataset was assigned a score that indicates how useful it is. The most important feature of the model was assumed to be given a higher significance score. The relevance score of each feature was compared to the scores of the other features in the dataset to determine its importance. Based on the findings of the study, ulcer scar, cornea appearance, types of cattle, inappetence, lacrimation duration, supranuclear

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

Ethiopian International Journal of Engineering and Technology (EIJET)
Volume 1, Issue 2, 2023

perforation, corneal perforation, blindness, corneal abrasion, and depression were the top ten important features in the dataset for identifying pinkeye disease. Whereas, cattle history, corneal ulceration, and corneal edema were the least important features in the dataset. Various experiments were conducted using all attributes, the top 5 and top 10 attributes as depicted in Fig. 2. Due to the insignificant changes in the overall performance of the models, the results presented in this article were based on the full feature dataset.

IV. Results and Discussion

Various experiments were conducted with different train/test split ratios. Depending on the maximum amount of data the researchers found from the source, the study used a trial-and-error method to compare several percentage splits experimentally, including 60/40, 70/30, 80/20, and 90/10. Based on the results obtained from the trial-and-error experiments, a 70/30 split was selected, in which 70% of the pinkeye dataset was used for training and 30% was used for testing (unseen dataset). Additionally, the majority of researchers in machine learning with relatively the same amount of data use the 70/30 percentage split, which is the most popular. Therefore, based on the preliminary experiments and previous research experiences, the 70/30% train/test split ratio was selected. In addition, preliminary experiments and literature surveys have been conducted to select the most appropriate models. Based on this related literature review and previous research experience, the models Random Forest (RF), XGBoost (XGB), AdaBoost (AB), and Artificial Neural Network (ANN) were selected for further experiments. Due to this, further experiments were conducted on four prominent machine learning models RF, XGB, AB, and ANN with a 70/30% train/test split ratio. Furthermore, several performance evaluation methods such as accuracy, recall, precision, and F1-score were used to evaluate and compare the performances of the selected models. Table I shows the results of four machine-learning models using different evaluation metrics. It is also worth mentioning that the researchers employed a grid search method for model hyperparameter tuning. Usually, Grid search is used to find the optimal hyperparameter values for all the selected models for optimal model performance.

Table I: Summary of Evaluation Result

Evaluation Metrics	*Algorithms*			
	RF	XGB	AB	ANN
Correctly classified instances	1621	1637	1628	**1639**
Incorrectly classified instances	32	16	25	**14**
Precision	97.09	99.22	97.85	**99.68**

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

ISSN (E): 2959-3921
(P): 2959-393X
DOI: https://doi.org/10.59122/144CFC14
Ethiopian International Journal of Engineering and Technology (EIJET)
Volume 1, Issue 2, 2023

Recall	97.99	**98.30**	**98.30**	98.15
F1-Score	97.54	98.76	98.07	**98.91**
Accuracy (%)	98.06	99.03	98.48	**99.15**

The test results of the four selected models are presented in Table I. It shows that ANN outperforms the other models with an overall accuracy of 99.15% followed by XGB with an accuracy of 99.03%. Furthermore, ANN outperforms the rest of the models in terms of the other performance measurement methods applied except for recall. Thus, we can assert that the ANN model performs better in all evaluation metrics except for recall than the rest of the classifiers tested. The experiments show that **1639 (99.15%)** of the test datasets are correctly classified while **14 (0.84%)** of the test datasets are incorrectly classified. Furthermore, one way to visualize the performance of classification models in machine learning is by creating an ROC curve, which stands for the "receiver operating characteristic" curve. Often, we may want to fit several classification models to one dataset and create an ROC curve for each model to visualize which model performs best on the data. Fig. 3 shows the ROC curve plot for the Artificial Neural Network model.

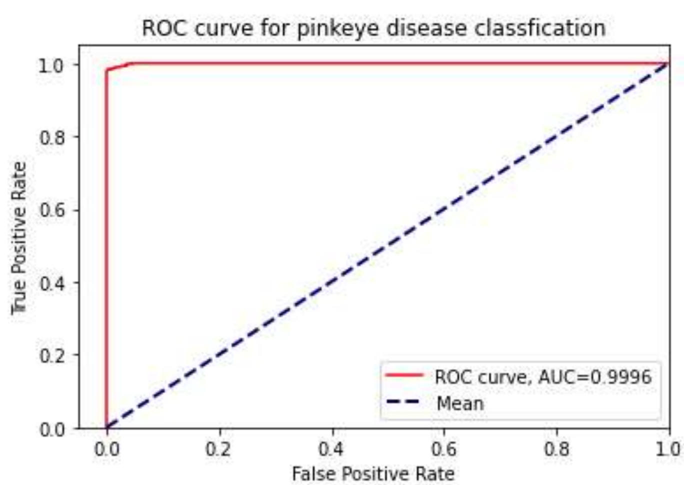

Fig. 3. ROC Curve Plot for ANN Classifier

Finally, based on the results of the feature importance analysis, the top ten determinant features in pinkeye disease classification include ulcer scar, cornea appearance, types of cattle, inappetence, lacrimation duration, supranuclear palsy, corneal perforation, blindness, corneal abrasion, and depression. While the least three important features include corneal edema, corneal ulceration, and cattle history. Further, the ANN model outperforms the other models with an overall accuracy, precision, recall, and f1-score of 99.15%, 99.68%, 98.15%, and 98.91%, respectively. Hence, the researchers attest that the ANN classification model is most suited to classify pinkeye disease.

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

V. Conclusion

The study aimed to build a classification model for cattle pinkeye disease by identifying relevant attributes. The initial data was gathered, converted from a manual to an electronic format, and then preprocessed so that it was suitable for analysis. The dataset contained 5508 records with 21 independent variables and a target class (Pinkeye and Not-Pinkeye). Essential procedures have been applied to realize an optimal pinkeye disease classification model. Based on the findings of the study, ulcer scar, cornea appearance, type of cattle, inappetence, and lacrimation duration are some of the major attributes for identifying the disease. Cattle history, corneal ulceration, and corneal edema do not have a significant effect on the model. Based on the experimental results the ANN classifier is found to be the best classification model with an overall accuracy of 99.15%, which is selected as an appropriate model to classify cattle pinkeye disease.

References

1. Feed the Future USAID, "Ethiopia's Livestock Systems: Overview and Areas of Inquiry," Gainesville, FL, USA: Feed the Future Innovation Lab for Livestock Systems (USAID), 2021.
2. A. A. Ali, C. J. O'Neill, P. C. Thomson, and H. N. Kadarmideen, "Genetic parameters of infectious bovine keratoconjunctivitis and its relationship with weight and parasite infestations in Australian tropical Bos *taurus* cattle," *Genet. Sel. Evol.*, vol. 44, no. 22, pp. 1 -10, 2012. https://doi.org/10.1186/1297-9686-44-22
3. N. Alias, F. N. M. Farid, W. M. Al-Rahmi, N. Yahaya, and Q. Al-Maatouk, "A modeling of animal diseases through using artificial neural network," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 3255–3262, 2018.
4. A. Seid, "Review on infectious bovine keratoconjunctivitis and its economic impacts in cattle," *J. Dairy Vet. Sci.*, vol. 9, no. 5, pp. 001-008, 2019.
5. L. Strickland, "Infectious bovine keratoconjunctivitis cattle pinkeye," *UT Ext. Inst. Agric.*, 2008.
6. M. Arnold, "Infectious bovine keratoconjunctivitis (pinkeye) in Cattle," *Vet. Clin. North Am. - Food Anim. Pract.*, vol. 31, no. 1, pp. 61–79, 2015.
7. X. Zhou, C. Xu, H. Wang, W. Xu, Z. Zhao, M. Chen, B. Jia, and B. Huang, "The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms," *Animals (Basel)*, vo.12, no.10, pp. 1 - 14, 2022.
8. K. P. Suresh, Dhemadri, R. Kurli, R. Dheeraj, and P. Roy, "Application of artificial intelligence for livestock disease prediction," *ACM SIGSOFT Softw. Eng. Notes*, vol. 13, no. 1, pp. 32–38, 1988.
9. D. Ashar, A. Kanojia, R. Parihar, and S. Kudoo, "Livestock disease prediction system," *VIVA-Tech*

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***

ISSN (E): 2959-3921 *Ethiopian International Journal of Engineering and Technology (EIJET)*
(P): 2959-393X *Volume 1, Issue 2, 2023*
DOI: *https://doi.org/10.59122/144CFC14*

Int. J. Res. Innov., vol. 1, no. 4, pp. 1–3, 2021.

10. H. P. P. Mon and O. Htwe, "Mobile-based cattle infectious disease prediction system," *Int. J. Sci. Res.*, vol. 7, no. 10, pp. 694–697, 2018.
11. M. Kneipp, A. C. Green, M. Govendir, M. Laurence, and N. K. Dhand, "Risk factors associated with pinkeye in Australian cattle," *Prev. Vet. Med.*, vol. 194, no. May, p. 105-432, 2021.
12. H. G. Olson, J. D. Loy, M. L. Clawson, E. L. Wynn, and M. M. Hille, "Genotype classification of Moraxella bovis using MALDI-TOF MS profiles," *Front. Microbiol.*, vol. 13, pp. 1–6, Dec. 2022. https://doi.org/10.3389/fmicb.2022.1057621

Received: *July 26, 2023;* ***Revised****: 17 September 2023;* ***Accepted:*** *04 October 2023;* ***Published:*** *31 December 2023.*
Corresponding author- ***Dr. Mohammed Abebe***