[1]Mr. Jayesh Surana
[2]Dr. Sanjay Kumar Sharma

Design a Deep Learning Model for Dairy Cow Abnormal Behaviour Prediction

Journal of Electrical Systems

Abstract: - When it comes to determining the health and well-being of cattle, it is critical to observe their behaviour. Another strong and cost-effective monitoring approach is a neural network-based monitoring system that analyses time series data from inertial sensors attached to cows. This is one of the most powerful and cost-effective monitoring systems currently available. Even though deep learning has made significant advances in pattern detection, large datasets are still required in many cases. When only a little amount of data is available, data augmentation is a very effective and low-cost pre-processing step for neural network-based systems. To study and analysis, a variety of ways for enhancing inertial sensor data based on cow behavioural data are discussed. Using convolutional neural networks, researchers are attempting to solve the difficult problem of identifying cow behaviour, which is made much more difficult by a scarcity of training data. Our proposed approaches areimproving the effectiveness of deep learning (CNN+LSTM) in the classification of cow behaviour while also lowering the total system costs associated with data collection and labelling.

Keywords: Convolutional Neural Networks, long short-term memory, Behaviour Prediction, Deep Learning, Dairy Cow.

I. INTRODUCTION

It is widely established that the well-being and health of the animals involved in milk production have a direct impact on the quality and quantity of the final products [1], as well as the amount of the final products. Cow behaviour monitoring has a substantial impact on the rise in dairy. The behaviour of the animal must be carefully observed in order to accurately establish the animal's present health status. The farmer's eye observations are taken into consideration in the majority of cattle status evaluation procedures;however, this is not necessarily the case in all cases. In a recent study, inertial sensors were used to monitor changes in activity levels, and algorithms were proposed for automating the evaluation process with the use of these sensors. A neural network-based system can use time series data from inertial sensors on dairy cows to detect problems in real time, and this can be done in real time. This is a powerful and effective solution that is available at a minimal cost. Although convolutional neural networks (CNNs) are inherently inefficient, they have proven to be incredibly effective in a wide range of hard pattern recognition applications despite their inefficiency. Despite the fact that CNNs consume a significant number of computational resources. In the case of deep learning, the availability of a large amount of data in a timely manner has contributed to the current level of success. Although it is possible to apply CNNs to issues when only a few tagged datasets are available, it becomes more challenging to do so. Collection and annotation of large amounts of behavioural data can be a time-consuming and complex activity because of the high cost of the data collection method used. It appears that applying CNNs to activity categorization based on small-scale behavioural data is a difficult task to achieve success. A possible solution to this problem has been proposed and has previously been tested. Even if it is a difficult effort, it is frequently only possible to supply new data that fits the necessary labelling criteria with the assistance of subject matter experts. Using retained labels to enrich data in particular areas may not be immediately obvious, and this is especially true in some situations. Scaling an original dataset using acceleration data may result in some labels being updated because the strength of their movements can distinguish some labels from others. Further complicating matters is the fact that tagged data is only available in limited quantities due to external activity interruptions, high individual variability, and noisy labelling. The methods for augmenting inertial sensor data that we present in this research have been shown to be

[1] er.jayeshsurana@gmail.com,

Department of Computer Science and Engineering Research Scholar, Oriental University, Indore. (M.P.), India.

sanjaysharmaemail@gmail.com

Department of Computer Science and Engineering, Research Supervisor, Oriental University Indore, (M.P.), India.

effective for the difficult task of identifying cow behaviour using low-cost data collection systems based on CNNs. The next section goes into greater depth about our humanitarian endeavours. We train CNNs to classify cattle in the field using data from small-scale behavioural trials to classify cattle in the field. All of these inertial sensor data improvement methods are intended to maintain the labelling information intact, so they can be used in place of the labelling information. Aside from that, we utilize a CNN-based behavioural model to compare the various strategies and identify which is the most effective.

II. RELATED WORK

Cattle behave in accord with their instincts, sensory awareness, and previous experience. It is considered instinctual behaviour when it is demonstrated by the cow in an unprompted manner. Sensory behaviours are those that are the result of anything that has been heard, seen, smelled, or touched in the environment.

G. Suseendranet al.[3]An approach to tracing cattle movements and forecasting where they will be found that makes use of the Markov decision process is currently being investigated (LPS-MDP). This technology is used to track the movement of livestock along a defined geographic area. This information is gathered from the cows' collars and utilised to determine their typical mobility pattern.

Z. Wang, C et al[4]Small-scale dairy farms are able to employ this method, whereas large-scale dairy farms are unable to do so. It is necessary to develop a cow activity monitoring and early warning system, which collects data on cow activity and notifies and locates the specific cows as they move about. As a result of this, labour costs can be lowered. Having data on dairy operations available in the system is critical for assisting in the long-term growth of dairy farms, which in turn aids in the improvement of production efficiency and yields while also delivering economic benefits.

S. Ma et al[5]They discovered that the rumen temperature may be consistently estimated with reasonable accuracy by using moving distance data as an additional input for the body temperature estimation algorithm. Long-term remote monitoring of cattle health and early problem detection are both possible with the recommended system's estimated body temperature.

S. Muthumanickam et al[6]With little or no human interaction, if any, the planned study will involve the automatic measuring of cattle's body temperatures, according to the researchers. A user or the property owner is notified whenever an inconsistency is discovered in the readings. As compared to previous automated systems, this new initiative aims to address the limitations of power and cost in the automated systems. It is possible to eliminate the need for energy through the use of radio frequency identification technology.

Y. P. Pratama et al.,[7] As a result of machine learning, the collected information is kept and sorted into normal, moderately abnormal and abnormal health classifications. In addition, use the axis sensor to provide a graphical representation of the cows' position.

F. Başçiftçi et al[8]It was discovered in this study that Internet of Things (IoT) technology can be used to develop an effective technique of detecting and monitoring acidosis disease in cattle, as well as documenting nutrition and behavioural data. It is possible to wirelessly collect pH and temperature values of the rumen portion through a circuit in a laboratory setting where the circumstances of the rumen part are provided.

III. PROPOSED METHODOLOGY

Processing data before to usage is known as data preprocessing. In order to make the depth model as accurate as possible, it is not possible to enter all of the sensor data at the same time. The data must be segmented using a sliding window before inserting it into the model. A behaviour identification system proposed in this study is able to recognise both the primary action and the transition action at the same time. Because the transition action only lasts for a short period of time, selecting the correct window size is essential. If the window is excessively large, it is possible that important information will be lost. An increase in computing costs will occur if this is not done. In contrast to the two-dimensional image data collected by sensors, the one-dimensional time series created by data segmentation yields behavioural data. Before applying a deep learning model to a dataset, input data must be entered and modified. Following window segmentation, the data must be transformed into dimensions. An axes-by-axe matrix of sensor data is used to turn the data into a two-dimensional representation. Because it preserves the correlation between the sensors' axes, this data processing technique has an advantage over others. After that, the deep learning model is fed images that are similar to photographs. Data preparation Feature Learning Based

1-D-CNN model's model structure is shown in Figure 1. The uniaxial acceleration and gyroscope data are transformed into a two-dimensional image array, which is identical to the original data after dimensional translation. As part of the convolution neural network, the image features are input into the convolution layer, which is typically comprised of a convolution layer and a pooling layer The convolution layer uses a convolution kernel to apply convolution operations to the input image in order to achieve feature mapping. By sampling local features from the convolution layer's feature map, this layer can reduce the number of neurons and parameters needed for the final image. It is possible to automatically extract the action feature information from the original action data by stacking the convolution layer on top of the pooling layer [5]. The design of the CNN model has been outlined. CNNs are networks comprised of sequential convolution and pooling layers. In this specific model, there are three of each type, with a pooling layer following each convolution layer. These layers work together to generate several feature map representations, including those that capture action-specific characteristics.

A range of choices are available for the configuration of each convolution and pooling layer, as well as for the overall network settings. The convolution process itself involves applying two-dimensional filters (kernels) that analyze overlapping sections of the input data, which is constructed by stacking multiple consecutive frames. Notably, the three convolution layers within the model utilize 18, 36, and 72 kernels respectively, enabling the progressive extraction of increasingly complex action features.

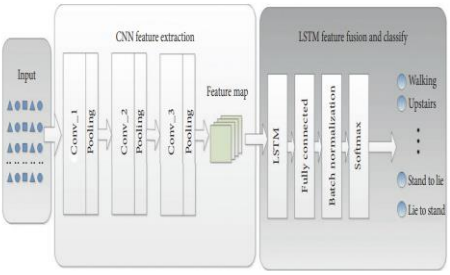

Figure 1: Cattle activity recognition framework based on CNN-LSTM

During the filtering process in a convolution layer, there's a potential issue: filters might not handle data at the edges of the input image effectively. To address this, a concept called "SAME" padding is introduced. This padding involves adding a border of zeros around the edges of the input image. This extra padding ensures that all data points within the image, including those on the edges, get a chance to interact with the filter. Consequently, the size of the output from the convolution layer remains the same as the input size, preventing unwanted data loss at the edges. Following the convolution operation, the resulting data (often called a feature map) typically undergoes a nonlinear activation function. This function injects a layer of non-linearity into the network, allowing it to learn more complex patterns in the data. The output of the activation function then becomes the layer's final output, ready for further processing in the network. The Sigmoid function, the ReLU function, and the Tanh function are all examples of popular activation functions. Among them, the ReLU function has the ability to transform the negative value of the data extracted by CNN into 0, while the positive value of the data greater than 0 retains its original value. A positive value larger than zero can be more clearly stated by the recovered features once a nonlinear processing operation has been performed. The activation function of ReLU is utilised in the convolution layer of CNN. It is as follows:

$$f(x) = \max(0, x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases} \quad (1)$$

Further, we have

$$f'(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases} \quad (2)$$

Some researchers believe that pooling layers offer advantages in reducing the number of feature maps and network parameters. Among the various pooling techniques, maximum pooling and average pooling are particularly common. Recent theoretical studies and performance evaluations have highlighted the superiority of the maximum pooling approach. This technique is widely used in deep learning and has received significant research attention. Notably, some studies suggest that maximum pooling is especially effective for sensor-based human behavior recognition tasks. Therefore, this paper exclusively employs the maximum pooling strategy throughout its CNN architecture, capitalizing on its strengths for the specific task at hand.

Feature Fusion and Action Classification are two approaches of categorising actions. Feature Fusion is one approach, and Action Classification is another. Both approaches are useful when it comes to categorising different types of tasks. In order to increase the recognition of transition events, an after-CNN network is used to turn the feature sequence of images composed of original data into the feature sequence f1, f2, f3, f4..., which is then transformed back into the feature sequence f1. Using the input LSTM and the input LSTM storage unit as inputs, we may generate a string of characters starting with m1 and ending with mn that can be used in other applications. The number of characters in the string is represented by the symbol mn. A few examples of LSTM memory units include the input gate, forgetting gate, and output gate, which are just a few of the many distinct types of memory units that are available in LSTM. Although it is conceivable for gradient disappearance to occur during the back propagation of a traditional circular neural network, it is feasible to mitigate this problem by using memory units in conjunction with learning weights. Since LSTMs are capable of collecting both global and time-dependent attributes at the same time, they are the most effective method for defining time-dependent behaviours. This makes them excellent for time-dependent behaviour analysis. LSTM cells are responsible for controlling the inward flow of information in neurons, and they are found in the thalamus and brain. The LSTM cell is made up of three gates: a forgetting gate, an input gate, and an output gate. The forgetting gate is the first gate in the LSTM cell. In this particular scenario, it is also important to employ the Tanh function in order to compute the predicted value of the LSTM cell. In this way, you can determine how much information can be sent from one cell to another in a certain amount of time. We can see how to determine the probability value using the following equation which can then be used to calculate the maximum quantity of data that can travel through the gate: The output of a neuron at the most recent time is described by its weight function (wf), which is represented by its bias function (b), and the current input of a neuron is represented by the value of its. Second, the input gate is made up of an update gate and a Tanh layer, which is responsible for regulating the amount of information that may be transmitted into the currently active cell. illustrate the calculation technique. When the output of the input gate is equal to or greater than the current value of the cell, the cell and the input gate are both updated at the same time.

The process from sequence analysis to action classification involves utilizing a Long Short-Term Memory (LSTM) layer to process data and generate vectors representing time and action correlations. These vectors are then inputted into a fully connected layer, responsible for amalgamating features to capture global action characteristics. However, as neural network models deepen, training complexities escalate due to evolving statistical properties of data flow between layers, requiring adjustments in learning rates to stabilize output data distribution. To tackle this challenge, Batch Normalization (BN) is proposed, standardizing values within each LSTM layer to maintain mean and variance consistency despite parameter distribution changes, thereby mitigating issues like vanishing or exploding gradients and facilitating faster training through potentially higher learning rates. Algorithm 1 details the BN method, calculating and normalizing input data mean and variance using mini-batch techniques, and incorporating scaling and shift factors (c and γ) learned during backpropagation to enhance accuracy. With enhanced feature visibility from BN, the normalized data is inputted into the Softmax layer for feature extraction and classification within the time series data. The Softmax function computes the posterior probability of potential actions at the output layer, with the output gate determining predicted values for each unit, iteratively producing the overall model output.

$$\Gamma_u = \sigma\left(w_u * \left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_u\right),$$

$$\widetilde{C} = \tanh\left(w_c * \left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_c\right),$$

$$C_t = \Gamma_u * \widetilde{C}^{\langle t\rangle} + \Gamma_f * C^{\langle t-1\rangle},$$

$$\Gamma_o = \sigma\left(w_o * \left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_o\right),$$

$$a^{\langle t\rangle} = \Gamma_o * \tanh\left(C^{\langle t\rangle}\right).$$

Among the many common cow behaviours we'll examine are feeding, walking, salting, ruminating, and sleeping. This is especially true if data is captured at 25 cycles per second, which yields datasets that total 5.89 hours in duration. [16] collected data for five fundamental actions using a CNN model with long short-term memory (CNN-LSTM), which took around 63 hours total (drinking, ruminating, walking, standing, and lying). To test the LSTM model's performance, the active video in [17] was around 50 hours long. This can be seen in [8].

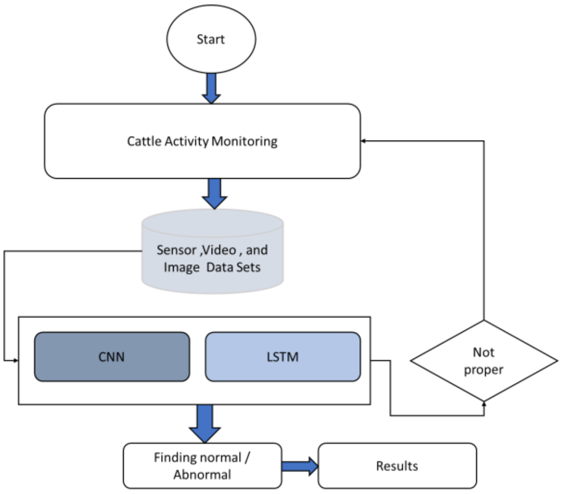

Figure 2:CNN model with long short-term memory (CNN-LSTM)

In the movie, which was approximately 30 hours in length, random rotation (ROT) based augmentation was demonstrated to supplement the data and deal with the uneven learning problem. For the purpose of creating a high-quality dataset, it is necessary to manually classify and annotate the data that has been collected, which is time consuming and labor-intensive. If deep learning is to perform effectively on small-scale datasets and completely realise its potential, it is critical that the processes used to augment behavioural data are accurate and effective. A variety of other activities, such as running, may create similar sensor data, such as acceleration data, which presents a dilemma in real-world applications because of the intricacy of animal behaviour. According to one hypothesis, similar hand motions are common among people working in a wide variety of industries, which could explain this peculiarity. Individual differences, in addition to interclass variability, exacerbate the situation. As a result, it is highly difficult to identify cow activity based on a small amount of data that is provided.

Results Analysis

Without the requirement for human involvement, an acceleration sensor was utilised to collect data on five cows in their natural surroundings without the use of any other equipment. An accelerometer was used to collect the data, which was sampled at a rate of 25 Hz. provides access to the whole dataset for the general public, which may be found at [3]. This letter is divided into three sections, each of which will be discussed in further depth below. The dataset is divided in half and then mixed with the learning and independent test sets in an 70/30 ratio to form a validation dataset, which is then divided in half again. Finally, ten percent of the learning set (equivalent to eight percent of the whole dataset) is utilised as a validation set to ensure that the model is correct. Training data is used to construct an initial behavioural model; the validation dataset is used to fine-tune model parameters during the training phase; and the independent test dataset is used to assess the effectiveness of the trained behavioural model.

To categorise cow behaviour, a CNN-based behavioural model and a range of data augmentation methodologies are utilised in conjunction with each other. The number of training instances in each scenario is depicted., and this number remains constant regardless of the circumstances. When analysing raw acceleration data that has not been supplemented with other information in order to obtain the baseline result, a CNN is employed. During the experiment, random parameter values are applied to the variables to see what happens. Every sequence of input data that is used in ROT is followed by the generation of a random ROT matrix. When determining whether or not to apply a strategy to a certain situation, REV employs a random selection procedure. Using a random selection technique, it is possible to determine the recombination ratio for mixed patterns. The ratio is formed by selecting a random number from the range of 0 to 1 and multiplying it by that number. Adding a time frame of less than 10 seconds to the input sequences allows the compensating technique to be executed successfully. A new piece of information is generated by altering the original training data, whereas a new piece of information is added to an existing piece of information by including more information into the original training data that would otherwise be rejected. It is not only the accuracy changes that are assessed, but also the aggregate augmentation outcomes of the other three ways when loss is taken into consideration that are examined.Top and bottom panels show the results of several data augmentation approaches for cattle activity classification; the top and bottom panels correspond to single and double data augmentation, respectively. On the left side of the figure, the average F1 score from each scenario is shown.

Table 1:Performance evaluation

Datasets Name	Precision	Recall	F1 Score
Ruminating	97.24	75.53	84.76
Resting	91.87	98.32	93.54
Salting	99.32	99.32	99.32
Feeding	99.55	99.55	99.55

The average F1 score for each scenario is depicted, with the numbers in the figure indicating the major findings for each scenario. When used in conjunction with other compensatory strategies, the proposed compensatory strategy outperforms a single data augmentation approach in terms of overall performance. With an average F1 score of 95.32 percent, Hybrid (CNN-LSTM)has the best classification performance of any of the four classes. Although problems in real-world applications may continue to exist, the strategies given here can assist in improving the outcomes of further investigations.

There are two non-deep-learning approaches that perform well in action classification recognition: random forest (RF) and Decision tree. Random forest (RF) and Decision tree are two examples of action classification recognition methods. As a result, the proposed CNN-LSTM model is evaluated in comparison to the RF and Decision treetechniques. Consequently To begin, feed the HAPT data set into the RF and Decision tree models. to say anything different, go back and segment the original

Table 2: Average accuracy of different activities with deep learning models.

Model	Accuracy
Decision tree	70.54%
Random forest	78.67%
CNN	91.45%
LSTM	89.33%
Hybrid (CNN-LSTM)	97,.73%

IV. CONCLUSION AND FUTURE WORK

It can be difficult to categorise the behaviour of cattle because of the small amount of data available In this study, a CNN-based behavioural model and the given data augmentation approaches successfully perform this demanding task. An increase in baseline accuracy of 89.33% to 97.73%is achieved by combining the enhancements, making for the best overall performance. This study revealed a low-cost method for keeping tabs on the well-being of farm animals and people alike. It will be decided which mode to use based on the present situation, and all data will be continuously fed into the thing speak account. Notifications will be made to the appropriate individuals and agencies in the event of an issue. Medications can be given to cows as soon as feasible if their health is abnormal. Thus, it is more convenient and helps increase the amount of milk produced.

REFERENCE

[1]. S. Kumar, M. Kumar Chaube and S. Kumar, "Secure and Sustainable Framework for Cattle Recognition Using Wireless Multimedia Networks and Machine Learning Techniques," in IEEE Transactions on Sustainable Computing, doi: 10.1109/TSUSC.2021.3123496.

[2]. G. Suseendran and D. Balaganesh, "Cattle Movement Monitoring and Location Prediction System Using Markov Decision Process and IoT Sensors," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), 2021, pp. 188-192, doi: 10.1109/ICIEM51511.2021.9445360.

[3]. Z. Wang, C. Cao, H. Yu and Y. Liu, "Design and Implementation of Early Warning System Based on Dairy Cattle Activity Detection," 2020 International Wireless Communications and Mobile Computing (IWCMC), 2020, pp. 2186-2189, doi: 10.1109/IWCMC48107.2020.9148122.

[4]. S. Ma et al., "Development of Noncontact Body Temperature Monitoring and Prediction System for Livestock Cattle," in IEEE Sensors Journal, vol. 21, no. 7, pp. 9367-9376, 1 April1, 2021, doi: 10.1109/JSEN.2021.3056112.

[5]. S. Muthumanickam, B. Keertana, L. Lavanya, S. Madhumitha and P. Sathvikaa, "Hardware Implementation of Health Monitoring of Domestic Animals Using RFID," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021, pp. 1-6, doi: 10.1109/ICSCAN53069.2021.9526431.

[6]. Y. P. Pratama et al., "Designing of a Smart Collar for Dairy Cow Behavior Monitoring with Application Monitoring in Microservices and Internet of Things-Based Systems," 2019 International Electronics Symposium (IES), 2019, pp. 527-533, doi: 10.1109/ELECSYM.2019.8901676.

[7]. F. Başçiftçi and K. A. Gündüz, "Identification of Acidosis Disease in Cattle Using IoT," 2019 4th International Conference on Computer Science and Engineering (UBMK), 2019, pp. 58-62, doi: 10.1109/UBMK.2019.8907035.

[8]. C. Li et al., "Data Augmentation for Inertial Sensor Data in CNNs for Cattle Behavior Classification," in IEEE Sensors Letters, vol. 5, no. 11, pp. 1-4, Nov. 2021, Art no. 7003104, doi: 10.1109/LSENS.2021.3119056.

[9]. Nleya, S.M., Ndlovu, S. (2021). Smart Dairy Farming Overview: Innovation, Algorithms and Challenges. In: Choudhury, A., Biswas, A., Singh, T.P., Ghosh, S.K. (eds) Smart Agriculture Automation Using Advanced Technologies. Transactions on Computer Systems and Networks. Springer, Singapore. https://doi.org/10.1007/978-981-16-6124-2_3

[10]. A. I. Mulla, A. P. Mulik, A. Prashant and D. D. Gawai, "Continuous health surveillance system for cattle," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 1192-1195, doi: 10.1109/ICCONS.2017.8250656.

[11]. M. Gjergji et al., "Deep Learning Techniques for Beef Cattle Body Weight Prediction," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9207624.

[12]. G. Suseendran and D. Balaganesh, "Cattle Movement Monitoring and Location Prediction System Using Markov Decision Process and IoT Sensors," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), 2021, pp. 188-192, doi: 10.1109/ICIEM51511.2021.9445360.

[13]. S. A. Schoenig et al., "Ambulatory instrumentation suitable for long-term monitoring of cattle health," The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004, pp. 2379-2382, doi: 10.1109/IEMBS.2004.1403689.

[14]. S. Jegadeesan and G. K. D. P. Venkatesan, "Distant biometry in cattle farm using wireless sensor networks," 2016 International Conference on Communication and Electronics Systems (ICCES), 2016, pp. 1-5, doi: 10.1109/CESYS.2016.7889964.

[15]. T. Natori, N. Ariyama, S. Tsuichihara, H. Takemura and N. Aikawa, "Study of Activity Collecting System for Grazing Cattle," 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), 2019, pp. 1-4, doi: 10.1109/ITC-CSCC.2019.8793451.

[16]. D. A. Widiasih, H. Susetya and R. Widayanti, "Early Detection of Leptospirosis in Cattle Urine Samples by Using Loop-Mediated Isothermal Amplification (LAMP) Method," 2018 1st International Conference on Bioinformatics, Biotechnology, and Biomedical Engineering - Bioinformatics and Biomedical Engineering, 2018, pp. 1-4, doi: 10.1109/BIOMIC.2018.8610637.

[17]. V. Gokul and S. Tadepalli, "Implementation of smart infrastructure and non-invasive wearable for real time tracking and early identification of diseases in cattle farming using IoT," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 469-476, doi: 10.1109/I-SMAC.2017.8058394.

[18]. H. Wang, A. O. Fapojuwo and R. J. Davies, "A Wireless Sensor Network for Feedlot Animal Health Monitoring," in IEEE Sensors Journal, vol. 16, no. 16, pp. 6433-6446, Aug.15, 2016, doi: 10.1109/JSEN.2016.2582438.

[19]. Surana, J. & Sharma, S. K. (2024). Predicting Cow Health with a Smart Framework: A Big Data and Deep Learning-Based IoT Approach. International Journal of Intelligent Systems and Applications in Engineering, 12(12s), 487–499.