Chapter 6

Deep Learning Framework for Recognition of Cattle Using Muzzle Point Image Pattern

6.1 Introduction

Recognizing animal repeatedly over time is a fundamental need for monitoring of livestock animals, the study of field-based ecology and related life sciences [127]. In scenarios, where photographic capture is achievable for recognition of different animals uniquely based on its biometrics characteristics. The discriminatory sets of biometrics feature of animals are extracted from the captured photograph or images using low-cost smart devices. The animal biometrics-based recognition system provides solutions for animal identification using non-invasive, cost-effective paradigms [99]. Therefore, the current research trends in the animal biometrics and computer vision are focused towards addressing the problems of identification of species, livestock management, and issues caused due to the failure of the classical animal identification systems and livestock based framework based system for recognition of individual cattle.

Towards successful operation of any farm, efficient management of herd, identification of animals, registration, and verification of false insurance claims of animals are necessary [119] [123]. Efficiency, affordability, and accuracy play the essential roles in the identification, verification of false insurance claims and monitoring of different cattle in modern farming [2] [166]. But the recognition of individual animals has been a serious problem in the traditional animal recognition system and livestock based frameworks for breeding associations, health management and verification of false insurance claims. Since classical animal recognition approaches (*e.g.*, ear-tagging based identification system, ear-tattoos based marking technique, embedded microchips, and freeze-branding based identification techniques) fail to provide the needed level of identification, registration, security of animal and monitoring of animals in traditional livestock framework based systems [9] [109].

Only a few types of research have been done so far to solve such major problems. Traditional identification approaches and non-biometrics techniques have such severe problems due to their boundaries and limitations for the identification, registration, and traceability of cattle using classical animal identification technologies. These technologies are not able to provide a competent level of security to animals.

On the other hand, identification of cattle has also been applied the texture feature based descriptor techniques on building the better representation of extracted feature of the face image and muzzle point images for robustness representation of local distortions, low illuminations, and poor image quality. However, most approaches are made of existing handcrafted features and appearance-based face recognition, and representation approaches for identification of cattle.

The handcrafted texture feature based representation approaches include the Circular-Local Binary Pattern(LBP) [3], Local Binary Pattern [142], Scale Invariant Feature Transform (SIFT) [192], Speeded Up Robust Features [19], and Gabor features [67]. Those general features are not designed primarily for the identification and verification tasks, and thus suffer from following significant problems: (1) some discriminatory characteristics of visual information can be lost in the feature extraction stage (mainly extracted features are the quantized phase), which unfortunately cannot be achieved in the later stages. Such prominent information loss rigorously destroys the performance of cattle recognition system, (2) the weakness of those handcrafted texture features is to need to be well aligned, which is considered to be considerably challenging to get or even not pragmatic for muzzle point images of cattle which has been taken in the unconstrained environment. These issues become even more challenging if different combinations of separate features, alignment techniques, and learning algorithms are examined for choice.

Recently, well-developed deep learning based machine learning methods enable to solve the biggest problems of classical animal identification based systems and livestock frameworks by learning the biometric feature representation of animals. The extracted features are classified by deep learning based classifiers jointly for identification and classification of cattle breeds in the computer vision tasks [61] [83] [106] [135] [174]. Thus, there is a need to develop an automatic recognition system for identification of cattle to solve the biggest problems of registration, heath monitoring of animals based on the unique identification of individual cattle. In this chapter, to address these problems cattle recognition system has been by using proposed deep learning based framework for the identification of cattle.

The proposed deep learning based recognition system performs the extraction of prominent set of texture features of muzzle point images. The resulting texture features of muzzle point images are classified by applying deep learning based Restricted Boltzmann machine (RBM) [112] framework for recognition of cattle. After pre-training of proposed recognition system, each Convolution Neural Network (CNN) [22] [106], Deep Belief Network (DBN) [81] [114], Stack Denoising Autoencoder (SDAE) [187] [188] technique

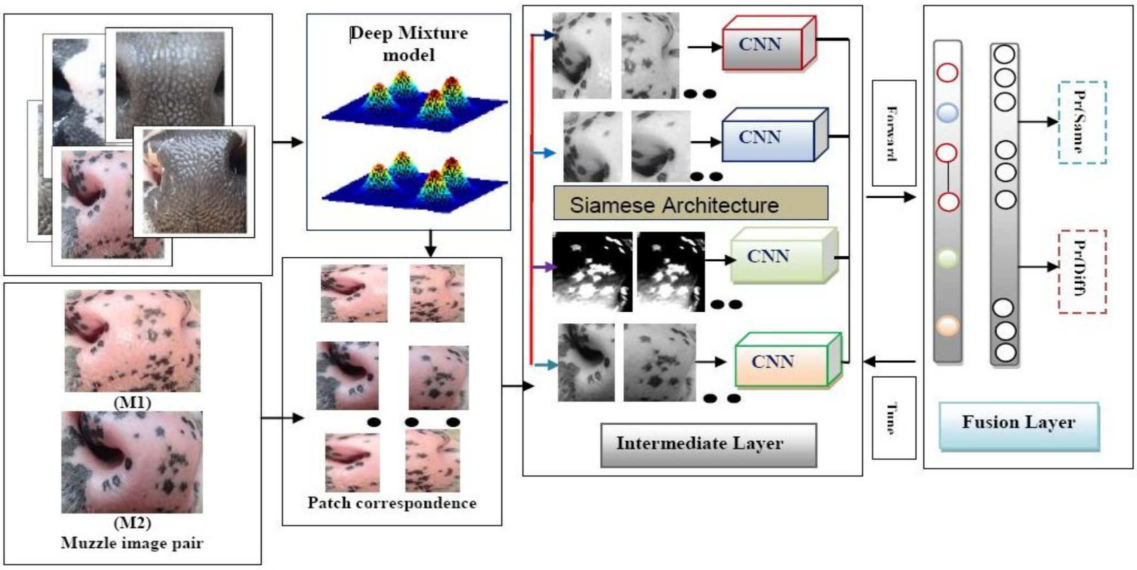

FIGURE 6.1: Proposed block diagram of cattle recognition using hybrid deep learning-based framework for recognition of cattle using muzzle point image.

have been applied separately, the entire proposed hybrid deep learning based approaches has been applied for representation of extracted texture features of muzzle point images to improve identification accuracy of proposed recognition system. The brief description of proposed cattle recognition system using hybrid deep learning based approaches is illustrated in next section.

6.2 Proposed System

In this section, the proposed Cattle Recognition System (CRS) is illustrated in detail for the recognition of individual cattle. The working of proposed recognition system is shown in FIGURE 6.1. The deep learning framework based representation algorithms are applied for learning the texture features of muzzle point images.

The working framework of the proposed cattle recognition system is in FIGURE (6.1). The proposed framework consists of the two-stage deep learning model to automatically learn robust local face representations for recognition of cattle. The proposed deep model automatically matches the local patches via deep mixture model (DMM) and then adopts convolutional neural network (CNN) to learn texture features for the better representation. Benefited from these two stages, the image representations of extracted features of muzzle point image are more robust to local variations regarding poor image quality, pose due to body dynamics and head movement, and also due to the poor illumination.

The acquisition of the spatial pixel values and appearance distribution over muzzle point images is shown in FIGURE (6.1). For each image pair, the pair of local patches of muzzle point images is obtained for each mixture component in DMM with regard to the corresponding responses. The selected patch pairs are then preprocessed with several illumination correction and enhancement methods and fed into multiple CNNs for supervised pre-training phase. The pre-trained sub-CNNs are finally fused together with a holistic fusion layer for better recognition of cattle.

The architectures of the hybrid convolutional neural network and DBN learning based framework are utilized for the encoding and decoding of texture features of muzzle point. It provides the better representation of variations and little illumination as well as poor image quality due to head movements and body dynamics of cattle (shown in FIGURE (6.2)).

The proposed approach is inspired from better representation using deep learning based feature extraction and representation techniques. In this chapter, convolution neural network [22] [106] and Deep Belief Network (DBN) [81] [114] learning techniques are applied to feature extraction and representation of muzzle point feature. The brief descriptions of applied deep learning approaches are illustrated in the subsequent subsections.

Convolution neural network (CNN) [22] is a deep-learning-based framework approach for the feature extraction and representation framework through which relations between the two muzzle point images regions are modeled hierarchically.

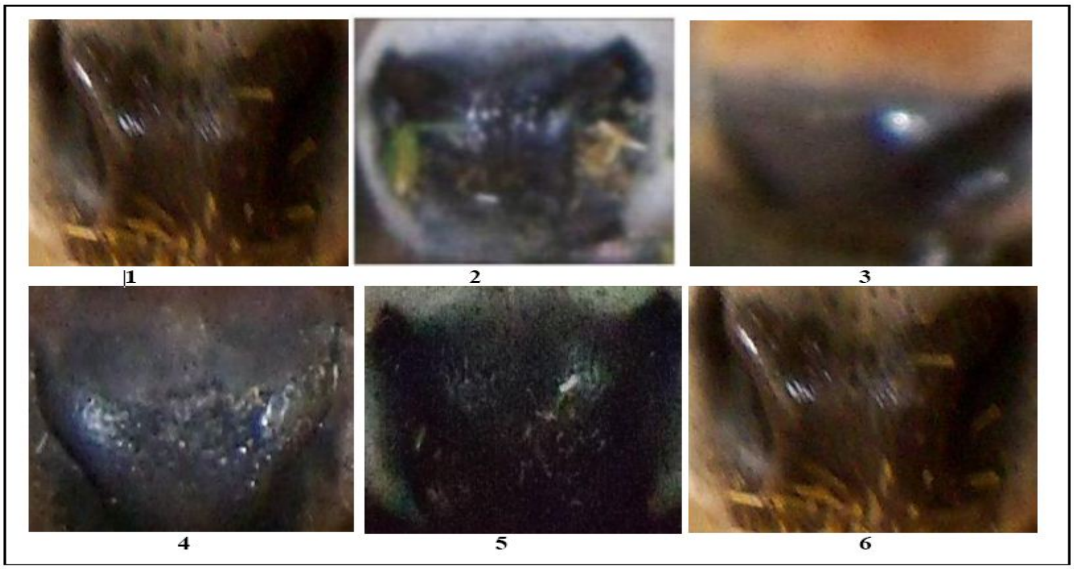

FIGURE 6.2: Illustrates the poor illumination, blurred images of muzzle point images from cattle database.

The CNN framework is built by stacking the multiple convolutional layers and max-pooling layers (shown in FIGURE 6.2, and FIGURE 6.3, respectively). The cascade of convolution layer and max-pooling layer based deep learning framework caters the robustness to shifting and rotation variations for the muzzle point images, captured in the unconstrained environment (*e.g.*, low illumination, body pose, and poor image quality) [106].

For the better analysis, the holistic face recognition and representation-based approaches [179], [60] [110] and the existing handcrafted texture feature based descriptor techniques [3] [142] have been used to perform the comparative analysis of experimental results.

In the proposed recognition system, first muzzle point image of cattle was taken to find the region of interest in the muzzle point images and extraction of features from the segmented image of muzzle point pattern. The extracted features of muzzle point images are mainly holistic features [179]. Compared with holistic features (pixel intensity value) [179], local texture features [142] are more robust to local distortions, low illumination, poor image quality and blurred images of muzzle point pattern of cattle. Therefore, the

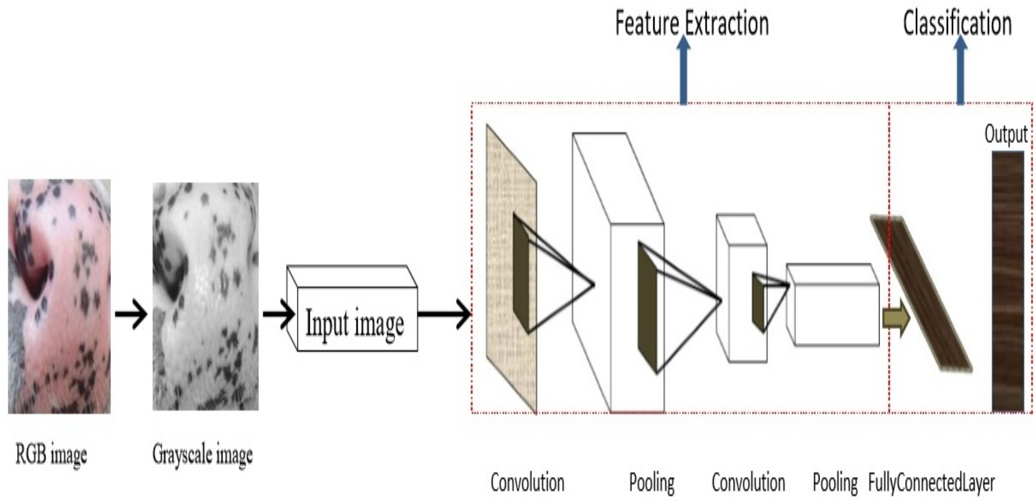

FIGURE 6.3: Illustrates the proposed convolutional neural network deep learning based framework for cattle recognition.

primary objective of this chapter is to design a hybrid deep learning model for capturing both holistic and local texture feature characteristics of muzzle point image database for identification of individual cattle.

Introducing local information to CNN technique enables the proposed architecture deep learning network (shown in FIGURE (6.1) and FIGURE (6.3)) to learn a more diverse and complex presentation and leads to potential improvement in the recognition accuracy of cattle. The mathematical formulation is illustrated for feature extraction and mapping of input muzzle point image pattern using convolution layer described in the next subsection.

6.2.1 Convolution Layer

The operation in each convolution layer is formulated as follows (shown in Equation (6.1)):

$$Y^j(r) = f(C^{j(R)}) + (\sum_i K^{ij(R)} \star X^{i(R)}) \tag{6.1}$$

In Equation (6.1), $(\star)$ denotes the convolution operation in the CNN deep learning approach. (X^i), and (Y^j) represent the (i^{th}) input map and the (j^{th}) output map in the CNN, respectively.

In the CNN network [61] [83] [106], K^{ij} shows the convolution layer kernel (*e.g.*, kernel based filter) for connecting the (i^{th}) input map and the (j^{th}) output map, respectively. Moreover, (C^j) presents the bias for the (j^{th}) output kernel map and $f(C^j)$ is a defined as a non-linear activation function for the proposed system. The non-linear activation function $f(C^j)$ is utilized element-wise for each neuron. The Softmax(0,.) non-linear activation function has been applied for better fitting abilities compared to traditionally used as non-linear activation function $(\tanh())$.

The term (R) depicts a local region in the defined layers where assigned weights are shared between the layers. Since muzzle point images are well defined structured as objects pattern, locally sharing weights of convolution kernels and biases in higher neural layers allows the network to learn different high-level features of muzzle point images in various locations.

In deep CNN based learning architecture network, a convolution layer can be illustrated $C(S,N,P,Q)$ where S, N, are the size of kernel of convolutional layer and total number of kernel maps. The various tuning parameters are defined such as (P) and (Q) for the selection of sharing of weights between the different layers in the deep CNN architecture [22].

In this chapter, it can be observed that locally shared weights between neurons plays a vital role in the higher convolutional kernel layers to enhance the fitting and generalization abilities of deep CNN based learning framework [22] [106]. Therefore, deep learning based convolution Neural Network (CNN) [22] [106] approach has been applied to extraction and representation the features of muzzle point images. In the representation, discriminatory features between the two images of muzzle point pattern are modeled hierarchically [22]. The mathematical formulation in the first convolutional layer of proposed system is shown for representation of muzzle point images as follows (shown in Equation (6.2)):

$$Y^j = f(C^j) + K^{(1,i)} \star X^1 + K^{(2,j)} \star X^2 \tag{6.2}$$

In Equation (6.2), (X^1) and (X^2) represent the two muzzle point image patterns (*e.g.*, ridge pattern) and these images are compared. During comparison, ridge pattern of muzzle images are convolved by the two convolutional kernels $K^{(1,i)}$ and $K^{(2,j)}$, respectively. Therefore, output function $Y^{(j)}$ presents a kind of similarity matching based scores (or relation) between the two muzzle point regions $X^{(1)}$ and $X^{(2)}$,respectively. The similarity match scores is determined by the two convolutional kernels $K^{(1,i)}$ and $K^{(2,j)}$, respectively.

The CNN networks apply a special deep learning architecture which is well-adapted to classify and recognize the muzzle point images of cattle. Using this deep learning architecture builds convolutional layers fast to train the proposed recognition system. This, in turns, helps us train deep, many layer networks, which are very good at classifying cattle based on muzzle point image database. The convolution neural networks based framework consists of three basic components: (1) local receptive fields, (2) shared weights and (3) pooling layer [106]. The brief description of each component is described in the next subsections.

6.2.1.1 Local Receptive Fields

In the fully-connected layers, the input muzzle images are depicted as a vertical line of neurons (shown in FIGURE (6.3)). In a convolution neural network, input as muzzle point images are taken as a 224×184 pixels, square of neurons, whose values correspond to the 224×184 pixel intensity value of used inputs muzzle point image pattern of cattle [22].

The proposed recognition system provides the input pixels of muzzle point images to a layer of hidden neurons. However, every input pixel of muzzle point images is not provided to every layer of hidden neuron. Definitely, we only make connections in small, localized regions of interest of the input muzzle image. To be more precise, each neuron in the first hidden layer is connected to a small region of the input neurons.

In this chapter, 5×5 is defined as filter size in the convolutional layer corresponding to 224×184 pixel input image of muzzle point pattern of cattle. The defined region in the each input muzzle image is called the local receptive field for the hidden neuron. The responses of local receptive field are generated in a little window size on the input pixels. Each connection between neuron learns locally sharing weights and hidden neuron learns an overall bias weight (C^j) as well to significantly improve the accuracy of muzzle point images. After that, responses of local receptive field are moved across the entire input muzzle image.

For each local receptive field, there is a different hidden neuron in the first hidden layer. Then we slide the local receptive field over by one pixel to the right (*i.e.*, by one neuron), to connect to a second hidden neuron and so on, building up the first hidden layer. In this chapter, 228×184 input muzzle point images with 5×5 local receptive fields are applied and then there can be 24×24 neurons in the hidden layer. This is because we can only move the local receptive field neurons across (23 neurons down), before colliding with the right-hand side (or bottom) of the input image.

6.2.1.2 Shared and Biases Weights

As mentioned in the above Equation (6.2), each hidden neuron has a bias weight (C^j) and filter size 5×5 weights connected to its local receptive fields [22] [106]. For training the proposed cattle recognition system, we have chosen the same 5×5 weight matrix and bias weight (C^j) for each of the 24×24 hidden neurons, shown as follows (shown in Equation (6.3)):

$$\sigma(b + \sum_{l=0}^{l=4} \sum_{m=0}^{m=4} w_{(l,m)} a_{(j+l),(k+m)}) \tag{6.3}$$

In Equation (6.3) for the (j^{th}) layers, output of (k^{th}) hidden neuron and (σ) is the neural activation sigmoid function, (C^j) is the shared weight value for the bias and $w_{(l,m)}$ is a (5×5) matrix of shared weights. The (x) and (y) are denoted the input activation at position (x, y).

For the recognition of image, more than one feature map are required to represent the extracted features in the feature space. The deep convolutional neural layer consists of several different feature maps.

In this chapter, three feature maps are applied in the convolutional neural network framework for mapping and better representation of extracted set of features of muzzle point images. Each feature map is defined by a set of 5×5 shared weights, and a single shared bias.

The result is that the network detects 3 different kinds of muzzle features, with each feature being detectable across the entire muzzle point image. The major advantages of sharing weights and biases are that it greatly mitigates the number of parameters involved in the convolution neural network [22] [106].

6.2.1.3 Pooling Layers

The first two convolution layers are followed by max-pooling layer for feature reduction and increasing their robustness to distortions of muzzle point images (shown in FIGURE (6.3)). The pooling layers are generally used immediately after convolution neural layers. The pooling layer has been applied to perform a down-sampling operation in the captured muzzle point images along the spatial dimensions (width, height), resulting in volume, such as [$224 \times 184 \times 3$].

The pooling layer takes each feature map output from the convolutional layer and prepares a condensed feature maps for uniform representation. The brief description of various types of pooling layers are given as follows:

1. **Max-pooling layer:** a pooling unit simply outputs the maximum activation in the 2×2 input region.

2. **Average pooling layer:**

   The average pooling layer performs down-sampling by dividing the input muzzle images into rectangular pooling regions 2×2. It computes the average of each region and returns the averages for the pooling regions.

3. **L2-pooling layer:** The L2-pooling layer performs the down sampling by dividing the input images and compute a pooling unit simply outputs the square root of the sum of the squares of the activation in the $2 \times 22 \times 2$ regions.

4. **Fully-Connected Layer:** Fully-connected layer computes the class scores, resulting in volume of size $224 \times 184 \times 3$, where the size of each muzzle point images are taken as 224×184 which correctly correspond to a given class score value.

5. **Rectified Linear Unit (ReLU):** ReLU layer applies an element wise activation function. The ReLU activation function is defined as $\max(0,x)$ thresholding at zero. This leaves the size of the volume unchanged ($32 \times 32 \times 12$).

In this chapter, a max-pooling layer is chosen to maximize the response of activation function of convolutional layer for muzzle point images as follows (shown in Equation (6.4)):

$$P(s) = Y^i_{(j,k)} = \max_{1 \leq p,q \leq s} (x^i_{(j-1)\times s+m(k-1)\times s+n}) \tag{6.4}$$

In the Equation (6.4), each neuron in the (i^{th}) output map (Y^i) pools over a 5non-overlapping local region in the (i^{th}) to map the input (x^i). The output of max pooling layer is represented by $P(s)$. In this chapter, 24×24 neurons output from the convolution layers, after max pooling operation 12×12 neurons are selected And, the network begins with 224×184 which is used to encode the pixel intensities of muzzle point images for the recognition of cattle [81]. This is then followed by a convolutional layer using 5×5 local receptive field and 3 feature maps. The muzzle point image has 224 wide, 184 high, 3 color channels), therefore a single fully-connected neuron in the first hidden layer of a CNN [22] [106] would have $224 \times 184 \times 3$ = 123648 weights (hidden feature neurons).

6.2.1.4 Fully Connection Layer

The final convolutional layer is followed by two successive fully-connected layers. The final layer of connections in the network is a fully-connected layer [162]. That is, this layer connects every neuron from the max-pooled layer to every one of the output neurons. The formulation of a fully-connected layer with neurons is represented by $f(N)$ with function as follows (shown in Equation (6.5)):

$$Y_j = f(\sum_{i=1}^{M} x_{(i)} \times w_{(i,j)} + C_{(j)}) \tag{6.5}$$

In the above Equation (6.5), M and N are total number of neurons of the current and previous layers, respectively.An element wise nonlinear activation function is chosen denoted by $f(N)$.

6.2.1.5 Rectified Linear Unit (ReLU)

Activation function in all encoder layers is set to rectified linear unit (ReLU) [113]. For supervised fine-tuning, a softmax layer with 20 nodes is added on top of each stacked auto encoder. the four dropout modules is chosen between layers to prevent overfitting. Dropout probability is set to 0.35 to remove the overfitting problem in this research work.

A ReLU function non-linearity based approach is applied for the first fully-connected layer (of the relational features) as did for the previous convolutional layers [113]. The hyperbolic tangent $(\tanh())$ non-linearity based activation function is applied for the output of single neuron to achieve a probabilistic based output for the matching of similarity of muzzle point image pattern of cattle [77].

Accordingly, the proposed deep learning model consists of two main layers: (1) the local layer and (2) the fusion layer. The local layer consists of various sub-CNNs corresponding to the local patches of muzzle point images and thus it captures the muzzle point features with regard to the local variations. The fusion layer contains a fully connected layer followed by a softmax classifier. It integrates the local responses to acquire a holistic view of the original image. Sub-CNNs are pertained separately to guarantee a certain level of independence. Such independence leads to a mutual complementary ability among sub-CNNs, resulting in considerable improvement with fusion layer.

6.2.1.6 Computation of Reconstruction Error

In this subsection, mathematical formulation has been applied to compute the feature sets in the proposed Convolution Neural Network (CNN)) base framework [106] [22]. Let us consider, $n \times n$ square neuron layer in the Convolutional Neural Network (CNN) framework is followed by convolutional layer. The $m \times m$ filter with weight has been applied to mitigate the noises from the muzzle point images using convolution layer and achieved the output of layer with size $(n-m+1) \times (n-m+1)$. The reconstruction of error is computed by evaluating the gradients of muzzle point images (pixel intensity) using gradient descent technique. First, errors are computed for the each previous layer by applying the partial differentiation techniques with respect to each response output of $\partial(Y^{l}_{(ij)})$. The partial differential has been used to compute the error which is denoted as $\frac{\partial E}{(Y^{l}_{(ij)})}$

For the computation of error of each layer, we find the gradient component for each weight $(w_{(ab)})$, where (a) and (b) are the neural layers and (w) are weight constraints between those layers. The chain rule of partial differentiation has been used to sum all the contributions of all terms in which the variable occurs (shown in Equation (6.6)) as follows:

$$\frac{\partial E}{(Y^{l}_{(ij)})} = \sum_{i=0}^{n-m} \sum_{j=0}^{n-m} \frac{\partial E}{(Y^{l}_{(ij)})} \frac{\partial (Y^{l}_{(ij)})}{(\partial w_{(ab)})} = \sum_{i=0}^{n-m} \sum_{j=0}^{n-m} \frac{\partial E}{(\partial x^{l}_{(ij)})} (Y^{(l-1)}_{(i+a)(j+b)}) \tag{6.6}$$

After calculation of reconstruction error between the input layer and output layer, the muzzle image features are represented for the recognition and classification of cattle. In this chapter, Restricted Boltzmann Machine (RBM) deep learning framework has been applied for the classification of extracted texture features of muzzle point images of cattle. The brief description of RBM is illustrated in the next subsection.

6.2.2 Restricted Boltzmann Machine

Restricted Boltzmann Machines (RBMs) [112] is deep learning approaches. It consists of a log-linear Markov Random Field (MRF) [136]. The RBM has undirected edges between the layers (shown in FIGURE (6.4)). The RBM deep learning framework is extremely applied for learning of extracted muzzle point features and representation of features from the given large unlabeled image database. The energy function in RBM framwork is linear in its free parameters. It is formulated as follow (shown in Equation (6.7)):

$$E(x,h) = -(a^T) - (b^T h) - (x^T W h) \tag{6.7}$$

In Equation (6.7), (x), (h) and (W) represent the visible input, hidden units, and weight matrix , respectively. Weights $(w_{(i,j)})$ signifies weight of connection between the hidden units (h_j) and visible unit (x_i) and represents the bias weights for visible units and (b) represents the bias weight for hidden layer. The probability distribution of a RBM [112], over the hidden and visible units is calculated as follows (Equation 6.8):

$$P(x,h) = \frac{1}{Z} exp(-E(x,h)) \tag{6.8}$$

In the above Equation (6.8) (Z) represent the partition function, is normalization constant. This further leads to the formulation of marginal probability which the sum of all the possible combinations of the hidden layers configurations given as follows (Equation (6.9)):

$$P(x) = \sum_h P(x,h) = \frac{1}{Z} exp(-E(x,h)) \tag{6.9}$$

6.2.3 Deep Belief Network

A deep belief network (DBNs) [81] [114] is graphical models which learn to extract a deep hierarchical representation of the training data. It can be form by stacking of RBMs (Restricted Boltzmann Machines) and trained in a greedy manner. This model applied the joint distribution between the input vector and the hidden layers, is a conditional probability for the visible units conditioned on the hidden units of the Restricted Boltzmann Machine (RBM) [112] at level, and is the visible hidden joint probability in the top-level RBM (shown in FIGURE (6.4)). The principle of greedy layer-wise unsupervised training can be applied to DBNs [114] with RBMs[112] as the building blocks for each layer. The DBN deep learning algorithm is given as follows (shown in Algorithm (3)):

Algorithm 3 Deep Belief network algorithm

procedure FUSION((s_1, s_2, s_3), (W_1, W_2, W_3))
    Initialization: Image data set $[X] = [X_1, X_2, \cdot, X_n]$ set of extracted pixel intensity values of images (N)
    Training phase:train the DBN deep learning model (M).
    Train the first layer as an RBM that models the raw input as its visible layer.
    Use the output of first layer as input for second layer.
    During representation of features in second layer gives two solutions:
    This representation can be chosen as being the mean activation $p(h^1 = h^0)$
    Train the second layer as an RBM as shown in FIGURE (6.4), taking the transformed data (samples or mean activation) as training examples (for the visible layer of that RBM).
    Iterate (3 and 5) for the desired number of layers, each time propagating upward either samples or mean values.
    Fine tune the parameters using supervised gradient descent.
    return $S_{(fused)}$

For the classification, we use the logistic regression classifier based on $h_{(l)}$ (last hidden layer of DBN model [114]); this step is similar with the using the weights and hidden layer biases, generated with unsupervised training, to initialize the weights of a multi-layer layer perceptron network [157]. The training of proposed deep model is shown in FIGURE (6.4). In this approach, CNN [106] [22] and DBN deep learning model [81]

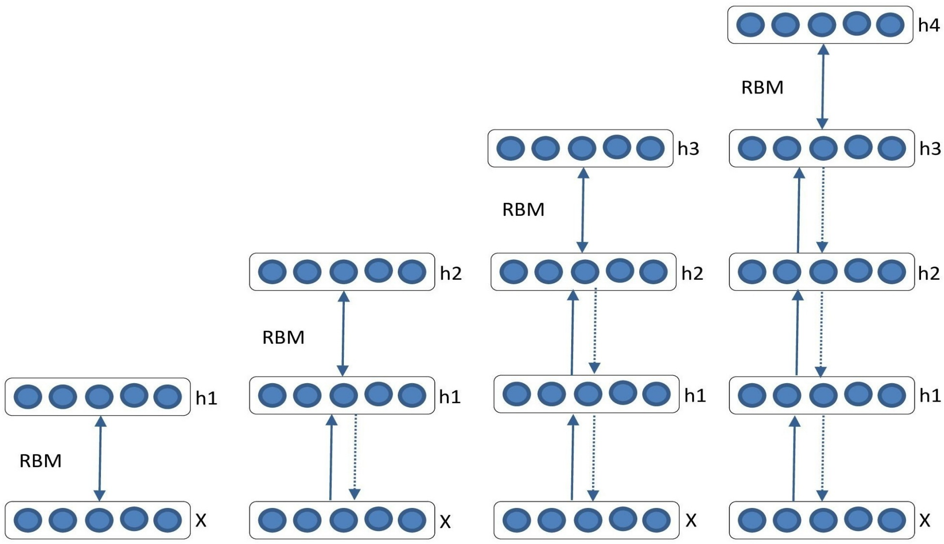

FIGURE 6.4: Illustrates the architecture of DBN composed of 4 stacked RBM deep learning approaches.

[114] is applied by constructing the multiple RBM models are stacked on top of each other. Each layer consists of multiple nodes which feed into the next layer.

6.2.4 Stacked Denoising Autoencoder

Stacked Denoising Autoencoder (SDAE) [187] is an extension of the stacked autoencoder. Stacking denoising autoencoders to initialize a deep network works in much the same way as stacking RBMs in deep belief networks (shown in FIGURE (6.4)).

Denoising autoencoders are arranged to form a deep network by nursing the latent representation (output code) of the denoising autoencoder found on the layer below as input to the current layer [187] [188]. The unsupervised pre-training of such architecture is done by one layer at a time. Each layer is trained as a denoising auto-encoder by reducing the error in reforming its input (which is the output code of the previous layer).

Stacking is done such that the output layer of the first autoencoder acts as the input layer of the second autoencoder. The autoencoder consists of two components: (1) encoder and

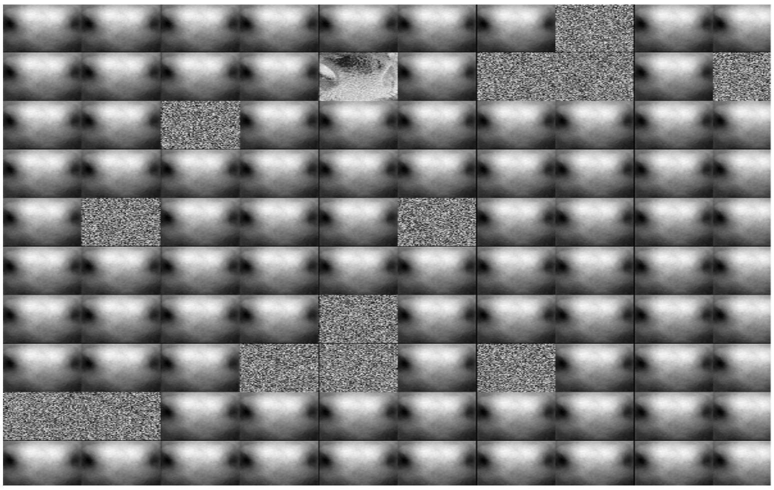

FIGURE 6.5: Illustrates the corrupted by (20%) of muzzle point images pattern using SDAE technique

(2) the decoder. The encoder performs the mechanism for transforming the input feature vector into a hidden representation while the decoder maps it back to the original input vector.

Once the first (K) layers are trained, after that (K+1) layers are trained and code or latent representation features are computed from the below layers. Specifically, the non-linear mapping function $f_{(w,a)}$ of the vectorized input image (X) is given as follows (shown in Equation (6.10)):

$$Y = f_{(w,b)}(X) = S(WX + b) \tag{6.10}$$

In the above Equation (6.10), $S(.)$ denotes the sigmoid activation function. W and b are the constraints of the mapping function. Next, in a similar manner, a reconstruction step $(G_{(w,b)})$ is implemented on the lower dimensional mapping (Y) as follows (shown in Equation (6.11)):

$$X^{"} = G_{(w,b)}(Y) = S(WX^{'}.Y + b) \tag{6.11}$$

Typically, $(X^{''})$ is an approximate reconstruction of input muzzle point image (X) and the SDAE autoencoder learns these features to minimize the reconstruction error between (X) and $(X^{''})$ by using backpropagation technique by greedily minimizing the loss function $(J_{(w,b)})$ as follows (shown in Equation (6.12), Equation (6.13) and Equation (6.14)):

$$J_{(w,b)}(x,X^{''}) = \|(X - X^{''})\|_2^F \tag{6.12}$$

$$minimum(\sum_{i=1}^{p})(\|(X - X^{''})\|_2^F) + \beta \sum_{i=1}^{m} KL(\rho, Y_i) \tag{6.13}$$

$$minimum\left(\sum_{i=1}^{p}(\big|\big|X - X^{''}\big|\big|_2^F) + \beta \sum_{i=1}^{m}(KL(\rho, Y_i))\right) \tag{6.14}$$

Where (ρ) is the size of input image of muzzle point pattern (m) and (β) are parameters used in the hidden layer of proposed deep learning architecture, (ρ) is weight for sparsity penalty term. The kullback-Leibler (KL) diveragence metric method [6] ha been applied to calculate the optimal sparsity penalty between two probability distributions (ρ) and $(\rho^{''})$, respectively. The kullback-Leibler divergence technique [6] has been defined as follows (shown in Equation (6.15))):

$$KL(\rho|\rho_j^{''}) = \rho log\frac{\rho}{\rho_j^{''}} + (1-\rho)log\frac{1-\rho}{1-\rho_j^{''}} \tag{6.15}$$

The (pre)-training of the SDAE based autoencoder [187] [188] is performed greedily on one layer at a time (*i.e.*, each layer of an auto encoder is trained by lessening the re-building of its input). Once all the layers are pre-trained successfully, the proposed deep learning based framework enters through a second phase of training, called fine-tuning

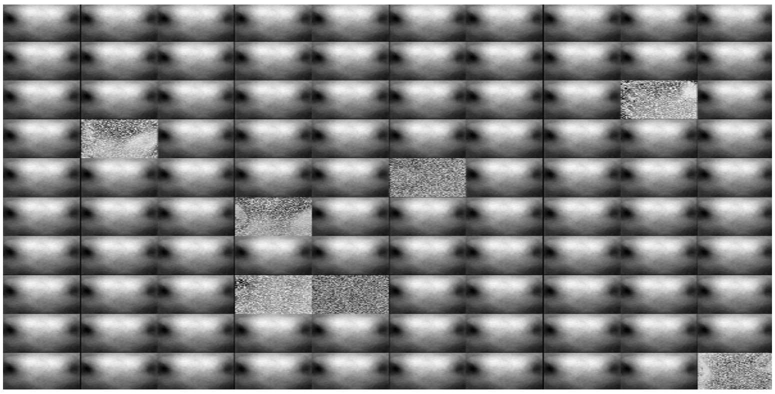

FIGURE 6.6: Illustrates the corrupted by (30%) of muzzle point image pattern using SDAE technique

using supervised gradient descent fine-tuning technique to diminish the prediction error on a supervised task.

Logistic regression layer has been included on the top of the proposed deep learning based framework (more precisely on the output code of the output layer). The entire deep learning network of proposed framework is trained using multilayer layer perceptron technique [157]. At this point, the encoding parts of muzzle point image feature pattern has been utilized by applying the SDAE autoencoder technique [187].

For the encoding the extracted muzzle features, unlabelled classes of muzzle point images are defined previously for a denoising autoencoders.

In this chapter, stacked denoising autoencoder technique [187] [188] has been applied for encoding the muzzle features for better representation in the feature space. The stacked denoising auto-encoder technique consists of phases for the encoding the features, namely (1) autoencoders [187] applied to encode the feature of input images for the training of the features and Multilayer Perceptron (MLP) [157] techniques are used as a logistic regression classifier in the proposed approach.

In the proposed approach, initially the input muzzle point images are transformed using a learnt non-linear transformation (ϕ). The transformation projects the input data into

Algorithm 4 Stacked Denoising Autoencoder (SDAE)

1: **procedure** ENCODING($(X_1, \cdots, X_n)$, $(Y_1, \cdots, Y_m)$)
2: (Initialize:) Image data set $[X] = [X_1, \cdots, X_N]$, $[Y] = [Y_1, \cdots, Y_N]$
3: Muzzle point images $M_{(X,Y)} \in$ N, N $\in$ size of muzzle database
4: **Initialize learning model**: SDAE model (M_1), Logistic model (M_2), and MLP model (M_3)
5: **Training phase:** train the SDAE model (M_1)
6: Train the first layer of proposed deep learning framework using the unsupervised techniques
7: Apply the models (M_1),(M_2) and (M_3) to train first layer $X = h(0)$
8: Use the output of 1^{st} layer as input for $2^n d$ layer
9: 2^{nd} layer represents the features using mean activation $p(h^1 = h^0)$
10: Train the 2^{nd}layer using the same techniques
11: Output of the 2^{nd} layer is fed as input to the 3^{rd} layer
12: Repeat step (9 and 10) to train the (K) number layer of proposed network
13: Apply the output of layers (step 10), to learn upward layers using mean values.
14: Compute the output from (K^{th}) layer
15: Multilayer perceptron model (M_3) is used to tune the network.
16: Fine tune the parameters (K^{th}) layers using supervised gradient descent.
17: Output of step (13) is fed as input data to logistic regression layer of model.
18: **Fine-tuning:** tune the whole trained network with the Logistic model on top of the layers
19: Apply the unsupervised training model to train and learn the trained model
20: **Output:** Encoded feature of muzzle point images form each autoencoders

a space where it becomes linearly separable. This intermediate layer is referred to as a hidden layer.

During pre-training, we used the first phase as trained model as a list of autoencoders [188]. The each stacked denoising autoencoder are trained separately to encode the muzzle point features of unlabelled muzzle point image database. In this chapter, the multilayer perceptron method [157] classification technique has been applied for training the deep learning model using transformation technique which is mentioned previously. The autoencoders and the sigmoid layers of the MLP [157] shared parameters. The latent representations computed by intermediate layers of the MLP are fed as input to the autoencoders (shown in FIGURE (6.5) and FIGURE (6.6), respectively). The algorithm of SDAE feature encoding and learning is shown in Algorithm (4).

6.3 Experimental Results and Discussion

In this section, the experiment results have been performed to compute the effectiveness of the proposed deep learning approach for the recognition of cattle using muzzle point image pattern. The comparison with existing benchmark algorithms (texture feature descriptor technique, appearance based feature extraction and representation, and learnt feature techniques) is accomplished to evaluate the identification accuracy in multiple identification settings.

For performance evaluation of experimental results, the database of muzzle point image pattern is segmented into following phases: (1) training phase, and (2) testing phase. In the training phase, 100 muzzle point images (10 cattle (subject) $\times$ 10 image of each subject) are utilized to train the proposed deep learning approach. In the testing phase, 400 testing pairs (40 cattle (subject $\times$ 10 image of each subject)) of muzzle point image pattern in which each folders are used to test the probe images.

The deep learning based framework demands the massive amount of database to train the proposed network. Despite, the size of muzzle point image database is 5000 images which are a relatively smaller image database. It is not satisfactory to adequately train a deep belief network or a stacked denoising autoencoder. Therefore, a transfer learning approach is applied for the fine tuning the weight between the input and hidden layer and determined the pre-training the proposed deep learning approach.

6.3.1 Algorithms For Performance Evaluation

The well-known existing handcrafted texture feature descriptor techniques have been applied for the evaluation of experimental results. The algorithms for performance evaluation includes the appearance based feature extraction and representation methods with

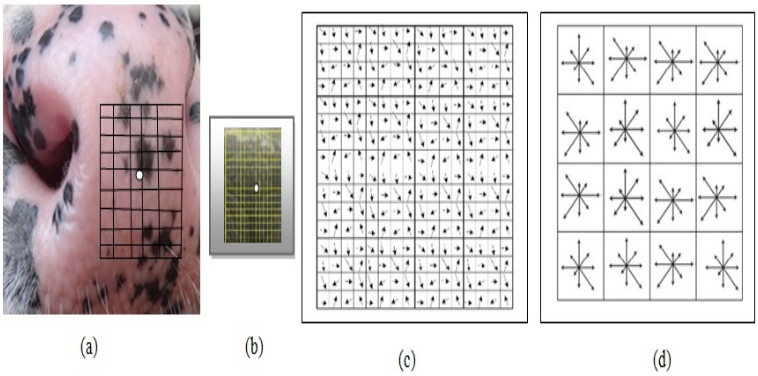

FIGURE 6.7: Illustrates the SIFT keypoint descriptor: (a) input muzzle point image (b) A SIFT descriptor of the size (400×400) is chosen from muzzle point image, (c) selection of 16×16 pixel orientations, (d) 4×4 cells descriptor with 8 pixel orientations are chosen. The size of single SIFT keypoint descriptor is $4 \times 4 \times 8$ = 128 element

learned feature based descriptor techniques for the evaluation and comparison of the experimental results of proposed approach. The handcrafted texture descriptor features and learnt features techniques are used for feature extraction and representation of muzzle point images of cattle. These handcrafted texture feature based learning approaches are mainly Local Binary Pattern (LBP)[142], Circular-LBP [3], Scale Invariant Feature Transform (SIFT) [120], Dense-SIFT [192], Speeded Up Robust Feature (SURF)[19], and Vector of Locally Aggregated Descriptors (VLAD) techniques [93]. The computation of features and encoding of the local binary pattern based feature, SIFT [120], and SURF [19] feature of muzzle point image is shown in FIGURE (6.7), (6.8), (6.9), (6.11), and (6.10), respectively.

6.3.2 Experiment Protocol and Experimental Evaluation

For the evaluation of performance, we have performed the three experiments to evaluate the effectiveness of the proposed approach and compare with existing benchmark texture descriptor technique and appearance based feature extraction and representation algorithms in multiple identification settings. In this subsection, the detail evaluation of

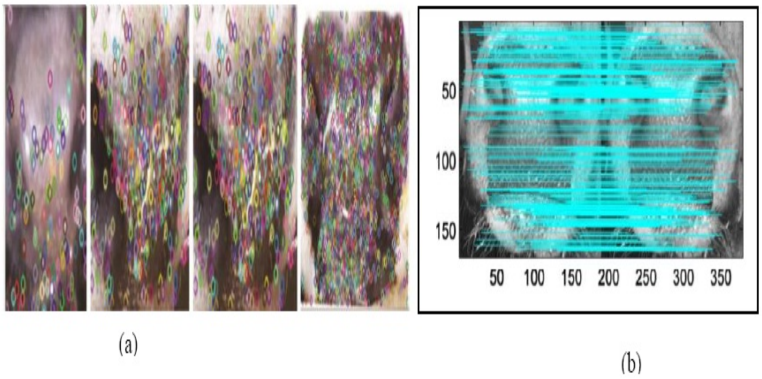

FIGURE 6.8: Illustrates (a) the process of SIFT keypoint localization; detection and (b) matching of test muzzle point image with stored muzzle point image using SIFT keypoint descriptor

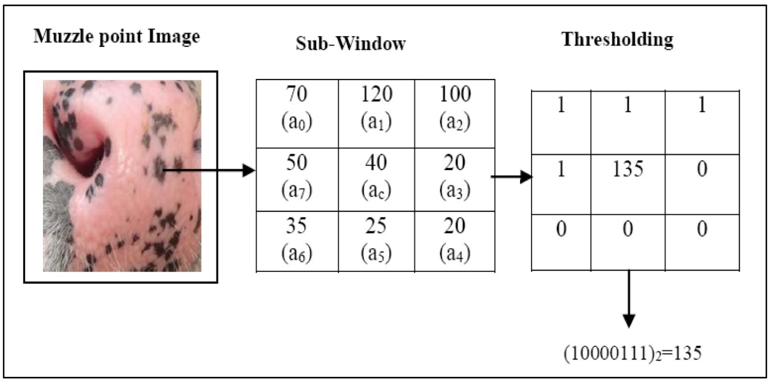

FIGURE 6.9: Illustrates the extraction and encoding of local binary pattern based descriptor features from the muzzle point image pattern

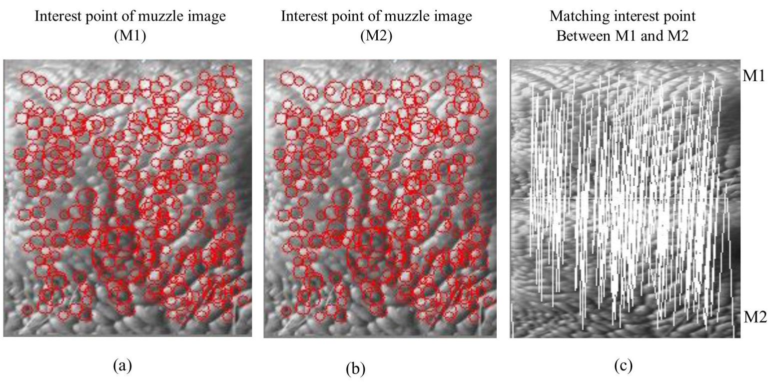

FIGURE 6.10: Illustrates the process of building of SURF descriptor: ((a)–(b)) shows the detection of keypoints in the muzzle point images, (c) matching of muzzle point images based on keypoint SURF descriptor with size (for a neighborhood of size 6s where s is scaling parameter of wavelet responses in horizontal and vertical directions.

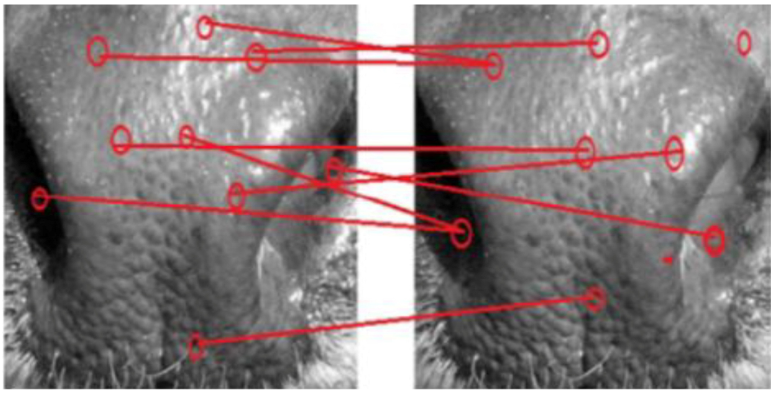

FIGURE 6.11: Illustrates the process of SURF matching of keypoints of muzzle point images using SURF descriptor based technique

experimental results is done for identification of individual cattle. Following experimentation has been performed to identification of cattle and validate the identification accuracy based on different identification settings:

1. **Experiment 1-Proposed Method**

   Muzzle point images of 100 cattle (subject) are randomly chosen for training the system and the remaining muzzle point images corresponding to 400 (cattle) are used for testing with 1, 2, 3, and 4 images per subject in the gallery. The least one shot similarity based matching scores are obtained per subject. It is used as the similarity match score. Further, all the experimental results are reported with five-times random sub sampling based cross validation technique. The identification experiments are performed and the results are reported in terms of rank-1 identification accuracy along with cumulative match characteristics (CMC) curve in FIGURE (6.12) and FIGURE (6.13), respectively.

2. The proposed approach has two components: (1) learning a robust muzzle point feature representation and learning the distance metric with one-shot similarity (OSS) [196] based matching technique. For evaluation of effectiveness of both the components, it is tactically replaced one component at a time with existing descriptors

TABLE 6.1: Illustrates the identification accuracy (%) of CNN, SDAE, and DBN deep learning approaches

S.No.	Proposed approach	Identification accuracy (%)
1	CNN	75.98%
2	SDAE	88.76%
3	DBN	95.99%

or matchers (chi-square (χ^2)) matching technique) and compared the results with four gallery images per subjects (cattle).

3. The experimental results are also analyzed with varying gallery sizes and summarized in Table 6.1, Table 6.2, and Table 6.3, and illustrated in FIGURE (6.12) and FIGURE (6.13), respectively. With the proposed deep learning approaches, such as CNN, SDAE and DBN yield 75.98%, 88.46%, and 95.99% identification accuracy, respectively.

4. In the Table 6.1, it can be noticed that with increasing the number of images per subject from one to four, the identification performance of DBN provides the highest identification accuracy. Table 6.1 illustrates the average identification accuracy based on individual patches of muzzle images for recognizing individual cattle.

5. Identification accuracies using the proposed deep learning framework are calculated by considering different number of muzzle point features amongst the calculated ones. It is observed that with the increase in usage of number of features the identification accuracies are gradually increasing in all the three deep learning algorithms (shown in Table 6.3). This explains the importance of all the features.

6. As explained earlier, as the number of features are increasing, the identification accuracies gradually increases but at every instant DBN is giving the best results amongst the three. This can be seen in FIGURE (6.12), and FIGURE (6.13), respectively.

TABLE 6.2: Identification accuracy (%) of muzzle point images of cattle based on CNN, SDAE, and DBN deep learning approach

Number of feature sets	CNN	SDAE	DBN
50	63.75%	67.75%	65.95%
100	67.98%	68.65%	69.85%
150	73.85%	71.96%	75.85%
200	76.75%	76.92%	77.94%
250	79.98%	78.67%	82.99%
300	82.99%	85.98%	86.92%
350	86.75%	89.79%	94.75%
400	92.98%	94.76%	98.99%

TABLE 6.3: Identification accuracy (%) of muzzle point images of cattle using CNN, SDAE and DBN deep learning approaches based on number of patches of muzzle point images

Number of patches	SDAE	CNN	DBN
1	62.89%	72.73%	78.43%
2	68.75%	77.86%	84.56%
3	72.87%	82.95%	89.91%
4	76.14%	88.03%	95.78%

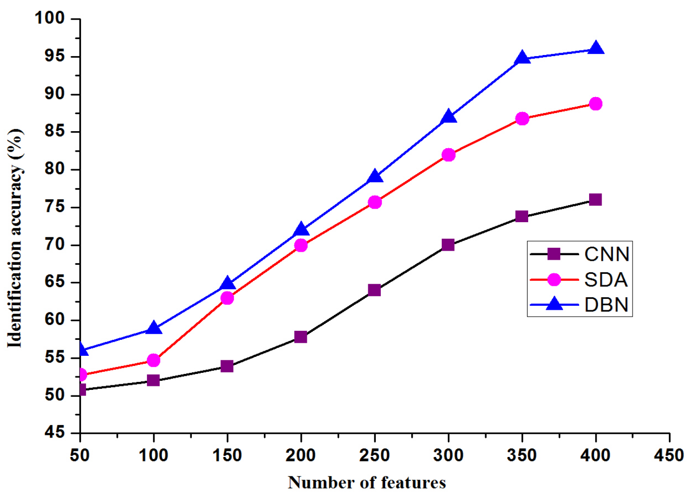

FIGURE 6.12: Illustrates the identification accuracy of CNN, SDAE, and DBN deep learning approaches based on extracted set of features.

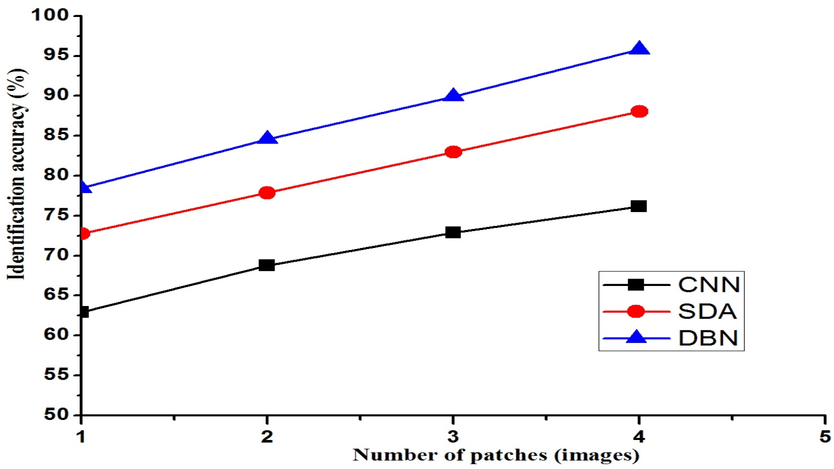

FIGURE 6.13: Illustrates the identification accuracy of CNN, SDAE, and DBN deep learning approaches based on extracted set of features ffrom number of patches (muzzle point images).

7. In the Table 6.2 and Table 6.3, identification accuracy of deep belief network is higher as compared to convolutional and stacked denoising autoencoder based approaches. The deep belief network deep learning approaches yields 98.99% of identification accuracy. It is observed that the accuracies of the three frameworks gradually increases with the increase in the number features of selected patches from the muzzle features. After the selection of 400 numbers of features from each patch (size: 200×200 pixels), deep learning algorithms provide the robust representation of muzzle features in the different layer of proposed framework. When size of each patches of images are reduced, selected discriminatory set of muzzle feature has also reduced, Therefore, we have selected the 400 number of features as better feature set from extracted patches. This shows us that all the patches calculated collectively describe the image to the best extent than a set of few patches.

TABLE 6.4: The experimental results are reported in terms of average accuracy with standard deviation over 10-fold cross-validation on holistic features based recognition algorithms

Number of muzzle point images per subject (cattle) in gallery				
Handcrafted Texture Features	1	2	3	4
LBP+(χ^2)[142]	72.84(0.8)	73.52(0.83)	73.76(1.23)	74.97(1.2)
Circular-LBP[3]	73.87(1.5)	76.50(1.83)	78.96(1.56)	79.87(2.2)
SIFT+(χ^2)[120]	67.98(1.2)	69.97(1.2)	70.97(2.2)	72.85(1.3)
DSIFT+(χ^2)[192]	69.75(2.6)	71.84(2.6)	73.98(1.4)	75.89(1.65)
SURF[19]	75.48(0.9)	78.99(1.5)	85.49(1.3)	89.76(1.43)

TABLE 6.5: The experimental results are reported in terms of average accuracy with standard deviation over 10-fold cross-validation for texture holistic features based recognition algorithms

Number of muzzle point images per subject (cattle) in gallery				
Existing Learnt Feature based method	1	2	3	4
VLAD+LDA+(OSS) [93][196]	45.98(1.5)	49.89(1.47)	53.94(1.22)	59.64(1.12)
VLAD+LDA+SVM [35] [93]	50.76(1.6)	54.92(1.27)	58.74(1.02)	67.98(1.17)

6.3.3 Experimental-2: Existing Handcrafted Texture Feature and Representation Method

In this chapter, the experimental results have been computed for the evaluation of performance measure of cattle recognition based their texture muzzle point image database using the existing handcrafted texture feature extraction and representation method. The experimental protocol used for the benchmark in identification settings for cattle recognition is same as experiment-1. 500 images (50 subjects (10% of the 5000 muzzle point images (*e.g.* 500 $\times$ 10 muzzle images of each subject (500 subjects (cattle)) from the muzzle point database of cattle are randomly chosen for training and the remaining muzzle images (corresponding to subjects) are used for testing phase.

Five times random sub-sampling is performed for validation of all the experiment results to ensure generality. The experimental results are also compared with appearance (holistic)-based feature extraction and representation [60] [110] [179], existing texture feature [3] [142] representation and metric learning algorithms. The hand-crafted features [142], learnt features, and appearance (holistic)-based feature extraction and representation [60] [110] [179] have been used for comparison are:

A. Hand-crafted features extraction based methods

- Local Binary Patterns (LBP)[142]
- Circular-LBP [3]
- Scale Invariant Feature Transform (SIFT) [120]
- Dense Scale Invariant Feature Transform [192] (DSIFT)
- Speeded Up Robust features (SURF)[19]

B. Existing Learnt feature based methods

- VLAD+LDA+OSS [93][196]
- VLAD+LDA+SVM+OSS[196]

C. Holistic or Appearance based methods

- Principal Component Analysis (PCA)[179]
- 2-D PCA [179]
- Linear Discriminant Analysis (LDA) [21]

- Kernel-LDA [103]
- Direct-LDA [121]

1. The existing handcrafted texture features based representation algorithms has been used for evaluations of experimental results. The experimental results are summarized in Table 6.4, respectively.

2. The existing handcrafted texture features algorithms such as Local Binary Pattern (LBP) [142] and Circular-LBP [3] feature descriptor based techniques provide the rank-1 identification accuracy of $72.84 \pm 0.80\%$ to $74.97 \pm 1.2\%$ and $73.87 \pm 1.5\%$ to $79.87 \pm 2.2\%$, respectively with 4 muzzle point images as gallery image per subject (cattle) (shown in Table 6.4).

3. The learnt feature SIFT feature descriptor techniques, such as SIFT [120] and Dense-SIFT (D-SIFT) [192] yield the identification accuracy of $69.75 \pm 2.6\%$ to $72.85 \pm 1.30\%$ and $69.75 \pm 2.6\%$ to $75.89 \pm 1.65\%$, respectively.

4. The evaluation of the experimental results, we performed testing of muzzle point images multiple in the gallery. During the experimental results on every test muzzle point images, we have calculated the least distance matching scores for every subject. The matching score values are used for evaluation of the experimental results.

5. The experimental results illustrate that even with multiple gallery muzzle point images, the identification accuracies are reported in the increasing order however, the performance leaning remains the same. The better performances of texture feature based descriptors algorithms are attributed to spatial collation in regional blocks that is able to good deal with the covariates, such as pose due to head movement, body dynamics, poor image quality and low illumination.

6. Vector of Locally Aggregated Descriptors (VLAD)[93][196] is a feature extractor learning based descriptor algorithm. It extracts visual information (features) from the training datasets [93][196]. As illustrated in given Table 6.5, the better recognition accuracy is achieved with OSS[196] (SVM-classification model [35]) for this representation when the gallery consists of four images per cattle subject, with rank-1 accuracy of $45.98 \pm 0.8\%$ to $59.64 \pm 1.12\%$ and hybrid VLAD+LDA+SVM [93] [35] [60] technique yields $50.76 \pm 1.6\%$ to $67.98 \pm 1.17\%$, respectively (shown in Table 6.5).

7. Based on overall performance of existing handcrafted feature descriptors techniques, SURF [19] technique provides the better identification accuracy of $75.48 \pm 0.9\%$ to $89.76 \pm 1.43\%$, respectively because the detection of the SURF [19] descriptor compute the discriminatory keypoints of muzzle point image pattern efficiently in the different scales (vertical and horizontal scale of wavelet responses) as compared to SIFT [120] and Dense-SIFT [192] approaches.

6.3.4 Experimental-3 Appearance based Face Recognition and Representation Method

Features of muzzle point images were extracted using the appearance based feature extraction and representation approaches, such as PCA [179], LDA [60] [21], and its modified version (*e.g.*, Kernel-LDA [103], Direct-LDA [121]) algorithms, respectively. The detail experimental results of appearance based face recognition and representation methods for cattle recognition is given as follows:

1. Principal Components Analysis (PCA) technique [179] is a useful statistical technique that has found application in fields such as face recognition. It is a common technique for finding patterns in data of high dimension.

TABLE 6.6: The experimental results are reported in terms of average accuracy with standard deviation over 10-fold cross-validation on holistic features based recognition algorithms

Number of muzzle point images per subject (cattle) in gallery				
Holistic Methods	1	2	3	4
PCA[179]	65.89(1.67)	68.95(1.83)	69.86(1.23)	70.97(1.20)
2D-PCA[78]	67.89(1.84)	69.65(1.63)	73.86(1.51)	75.67(1.45)
LDA [21]	69.96(1.98)	72.91(1.64)	73.48(1.43)	74.99(1.37)
Kernel-LDA [103]	60.89(1.87)	65.65(1.73)	67.86(1.57)	68.97(1.28)
Direct-LDA [121]	63.77(1.79)	65.96(1.63)	68.96(1.53)	69.97(1.33)

2. PCA technique is used to perform the dimensionality reduction of extracted feature of muzzle point image database. Principal components corresponding to 99% Eigen-energy in the PCA subspace are retained [179].

3. Recognition of cattle based on muzzle point pattern is performed using Linear Discriminant Analysis (LDA) technique [60] and Cosine similarity based matching approach [118].

   The Cosine similarity technique has been used to match a pair of muzzle point pattern to find the similarity value or score corresponding test muzzle point images. The identification accuracy are summarized and reported in Table 6.6, respectively. Based on overall observation, It can be observed that the Kernel-LDA [103] and Direct-LDA [121] recognition techniques provide the identification accuracy of $60.89 \pm 1.87\%$ to $68.97 \pm 1.28\%$,and $63.77 \pm 1.79\%$ to $69.97 \pm 1.33\%$ to recognize individual cattle based on muzzle point features.

On the other hand, texture feature descriptor techniques, such as SURF, LBP, Circular-LBP, SIFT, Dense-SIFT and VLAD are handcrafted existing features based descriptors for the better representation of muzzle point features for identification of individual cattle.

The LBP [3] [142] and Dense-SIFT [192] descriptor algorithm illustrates the low recognition accuracy as compared to VLAD+LDA [60] [93] with One-Shot-Similarity (OSS)

[196] and VLAD+LDA [93] with Support Vector Machine (SVM) [35] techniques for identification of individual cattle using their primary muzzle point image pattern. The learnt feature descriptor techniques, such as VLAD+LDA+OSS [60] [93] [196] and VLAD+LDA+SVM [35] [60] [93] approach yield $60.89 \pm 1.5\%$ to $59.64 \pm 1.12\%$ and $50.76 \pm 1.6\%$ to $67.98 \pm 1.17\%$ identification accuracy, respectively.

The capture muzzle point image database has major challenges due to the unconstrained environment, such as poor illumination, image quality, low contrast, and blurred images. Therefore, existing handcrafted feature extraction and representation approaches [142] are unable to perform the identification of individual cattle based on their muzzle point images. The learning-based feature extraction [93] and matching approaches [196] [118] cater an explicit encoding mechanism using stacked denoising autoencoder [188] [187] method of the extracted feature of muzzle point image to improve the recognition accuracy of individual cattle.

In this chapter, One-Shot Similarity (OSS) [196] matching technique using Fisher linear discriminant analysis (FLDA) [60] approach, and Incremental Support Vector Machine (I-SVM) classification model are applied to perform the classification of extracted features for cattle recognition. To validate the experimental results one-class online Incremental-SVM (1-online I-SVM) [172] model is used to classify the extracted texture features of muzzle point images of cattle. The main advantage of OSS similarity learning technique [196] to learn the deep learning model based framework using semi-supervised [187] based matching similarity technique for the classification and recognition of individual cattle. The OSS similarity learning technique selects unlabeled training muzzle point images as the set of negative constraints against two input images of muzzle point pattern are matched.

Based on the overall experimentation, and achieved recognition accuracy, it can be noticed that existing handcrafted texture feature extraction and representation algorithms are bound and have limited representation capacity of muzzle point image in feature space. Therefore, texture feature based descriptors are not applicable for better recognition of cattle in the specific problem domain.

The proposed deep learning based identification framework performs the encoding of the discriminant feature of muzzle point images for representation in the feature space [187] [188]. The machine learning based distance metric technique has been applied to learn the semantic representation of extracted features for understanding of the encoding scheme by using SDAE encoding technique [187].

Based on complete observations made and the challenges of cattle recognition using prominent feature of muzzle point images, we postulated that existing animal recognition methodologies and automatic recognition algorithms for species or individual animal that are tailored specifically for recognition of animal, via unambiguous training process. The deep learning based learning framework can be applied to perform the recognition and verification of individual cattle more effectively. The proposed deep learning based recognition system of cattle caters a friendly, non-invasive, robust as well as cost-effective solution for the recognition of animals.

6.4 Comparison to the state-of-the-art

In this chapter, detailed analysis of experimental analysis of proposed deep learning approach is done to evaluate the performances of cattle recognition based on muzzle point images. The performance comparisons of proposed deep learning framework based cattle system have been evaluated against the current state-of-the-art methods for recognition and verification of livestock animal based on their muzzle print images from the literature.

To perform the identification of cattle, the muzzle point image database is prepared from Department of Dairy and Husbandry, Institute of Agriculture Sciences (I.A.S), Banaras Hindu University (B.H.U), Varanasi, India using a 20-megapixel camera. For, capturing the muzzle image pattern of cattle, The manual acquisition methodologies are not applied using. In the classical acquisition methodologies of muzzle point images includes various equipment and materials (such as A5-size white papers, black ink and stamp (impacted), soft cottons, hard ropes, tissue paper and assistant team of the dairy staff members) to capture the images of muzzle print of cattle. The captured images of cattle on A-5 paper with blue ink have provided few motives to perform the research for identification and monitoring of cattle by researchers, scientists, and veterinary professionals. Therefore, captured muzzle point database of cattle using a 20-megapixel camera is original images of muzzle point pattern of cattle. The prepared database also includes various covariates of muzzle point images due to poor image quality, low illumination, pose, variation based images due to head movement and blurred.

In the available literature, there is no availability of muzzle point image database in the public domain. Very few researches have been done for the identification of cattle based on the muzzle print images. Based on printed muzzle images, we have compared the experimental results of proposed approach with previously published results on muzzle print images including results that use different computer vision and pattern recognition techniques. The comparative analyses of experimental results are shown in Table 6.7.

Noviyanto et al. [141] proposed a method using SURF and Eigenface based approaches for recognition of individual cattle. The major shortcoming of this paper is that the authors performed the experimental on the small dataset of muzzle print images. The proposed approach cannot handle various configuration of rotation and scale of muzzle print images. In [15], authors proposed a cattle recognition based framework for identification of cattle using Eigenvalues based approaches. The authors have applied principal component analysis techniques to mitigate the dimensionality of extracted features.

The proposed approach by Minagawa et al. [130] has not reported exactly experimental results the same due to the unexplained filtering techniques.

In [15], Barry et al. proposed cattle identification using principal component analysis and Euclidean distance classifier techniques based on muzzle print images. The proposed approach is selected for training separately on a different number of normalized muzzle images sets of 2, 4, 6, 8, and 10 training images from 29 cattle. The drawback of the approach is that an experimental result was taken on a separate set of 3 images only per animal. The authors have not performed any cross-validation of the experimental results. The implementation of the approach proposed by Barry et al. [15], is also not exactly the same due to the watershed segmentation technique is not implemented properly. The approach proposed by Barry et al. [15], has been very strict due to the false match has been zero and 468 false non-matches have been reported over 560 genuine matching.

In the similar direction, Awad et al. [10] proposed a method to improve the performance of classical animal identification system. For identification SIFT key-point matching based descriptor technique has been applied to compute the feature points of muzzle print images. For better identification, Random Sample Consensus (RANSAC) technique is utilized with the SIFT technique to mitigate the noises, outlier from captured print of muzzle images. The major limitations of this method includes: (1) no cross-validation test is done for validating the experimental results, (2) the pre-processing and enhancement techniques are not applied to improve the image quality of images, image outliers, blurriness, and poor image quality. To mitigate these major problems of cattle identification, the author Noviyanto et al. [140] proposed a matching refinement technique using a SIFT keypoints matching technique.

In [66], authors proposed a method for automatic recognition of cattle using Weber Local Descriptor (WLD) technique. The proposed approach extract the features from cattle muzzle print images (images from 31 head of livestock). The extracted features are classified by AdaBoost model [142] to identify the head of individual cattle from their WLD descriptor features. The limitation of this paper is that experimentations are carried out the small datasets, and no cross-validation technique is applied to validate the experimental results.

The authors Cai and Li [30] proposed a method for automatic recognition of cattle using the local binary pattern based feature descriptor and extended LBP descriptors techniques. The robust alignment by sparse and low-rank decomposition approaches is applied to align the face images of cattle due to poor illumination, image misalignment and occlusion problem in the test face image of cattle.

The dissimilarity between test and trained images are performed on a separate set of face images using the weighted Chi-square distance technique. The major shortcomings of this paper are that authors have not performed the experimental results on slight datasets, (2) no pre-processing steps are involved to reduce the noises of facial images, and finally (3) any cross-validation is not applied to verify the identification accuracy.

In the comparative study of the experimental results, the muzzle point image database of 100 cattle (subject) have randomly selected for training the proposed cattle recognition system, and the remaining muzzle point images corresponding to 400 (cattle) have been used for the testing phase. The testing phase includes 1, 2, 3, and 4 muzzle point images per subject (cattle) in the gallery. The least one shot similarity based similarity matching scores are accomplished per subject. Moreover, the five-times random sub-sampling based cross-validation technique has been implemented for validation of experimental results.

The proposed approach consists of two steps: (1) learning and representation of the muzzle point image -based texture feature representation and (2) distance metric with one-shot similarity (OSS) [196] based matching technique. For evaluation of the effectiveness of both the components, it is tactically replaced part of at a time with existing descriptors or matches (chi-square (χ^2)) matching technique) and comparisons are done to evaluate the experimental results with four gallery muzzle point images per subjects (cattle).

With the proposed deep learning approaches, such as CNN, SDAE, and DBN frameworks yield 75.98%, 88.46%, and 95.99% identification accuracy, respectively. In the Table 6.1, it can be observed that with increasing the number of muzzle point images per cattle from one to four images, the identification accuracy of DBN framework for cattle recognition caters the highest identification accuracy. Moreover, the average identification accuracy to identify the cattle based on individual patches of muzzle point image pattern gives better accuracy (shown in Table 6.1).

The comparison of experimental results is also evaluated using the existing handcrafted texture feature extraction and representation techniques for cattle identification based on their muzzle point images. The existing handcrafted texture feature descriptors: Local Binary Pattern (LBP) [142] and Circular-LBP [3] methods yield the rank-1 identification accuracy of $72.84 \pm 0.80\%$ to $74.97 \pm 1.2\%$ and $73.87 \pm 1.5\%$ to $79.87 \pm 2.2\%$, respectively with 4 muzzle point images as gallery image per cattle.

The learnt feature based descriptor methods includes SIFT [120] and Dense-SIFT [192] methods for identification of individual cattle. These methods yield the identification accuracy of $69.75 \pm 2.6\%$ to $72.85 \pm 1.30\%$ and $69.75 \pm 2.6\%$ to $75.89 \pm 1.65\%$, respectively. The evaluation of experimental results was also done on the muzzle point feature of cattle using Vector of Locally Aggregated Descriptor (VLAD)-based methods. It achieved the better accuracy with OSS [196] (SVM classification model [35]) for identification of

TABLE 6.7: Summary of cattle recognition and classification techniques based on muzzle images pattern used in the literatures and experimental results of proposed approach.

Comparison of our proposed approach with the literature			
Authors	Images	Technique used	Identification accuracy(%)
Awad et al.[10]	15 cattle	SIFT+RANSAC	93.30%
Barry et al.[15]	29 cattle	Eigen-value	98.50%
Cai and Li[30]	30 cattle	RASL+WLBP	95.30%
Gaber et al.[66]	31cattle	WLD+ABD	99%
Kumar et al.[109]	300 cattle	PCA+ LDA+ICA	85.95%
Minagawa et al.[130]	43 muzzle image	PCA+Eigen-values	30%
Noviyanto [140]	48 muzzle image	SIFT+PCA	0.0167 (EER)
Noviyanto et al[141]	80 muzzle image	SURF+Eigenface	89.30%
Tharwat et al.[177]	31 cattle	Gabor+SVM	99.50%
This res. study	**5000-muzzle point image**	**CNN,DBN,SDAE**	92.98,94.76,98.99**%**

cattle when the test image consists of four images per subject, with rank-1 recognition accuracy of 45.98 ± 0.8% to 59.64 ± 1.12%. On the other hand, hybrid VLAD+LDA+SVM [93] [60] technique yields 50.76 ± 1.6% to 67.98 ± 1.17% recognition accuracy for cattle recognition, respectively.

Based on overall observations, it can be concluded that proposed method uses the deep learning based framework for recognition of cattle. It achieves the significant improvements over all of this existing handcrafted feature descriptor, and baselines algorithms which are available in the literature.

In the Table 6.7, some notions are used, such as WLD = Weber's Local Descriptor ADB = AdaBoost classifier, WLBP = Weber's Local Binary Pattern Descriptor, RANSAC = RANdom Sample Consensus algorithm, CNN = Convolutional Neural Network, DBN = Deep Belief Network, SDAE = Stacked Denoising autoencoders, OSS = One-Shot-Similarity (OSS), SIFT = Scale Invariant Feature Transform, PCA = Principal Component

Analysis, LDA = Linear Discriminant Analysis EER = Equal Error Rate, SVM = Support Vector Machine, This res. study = This research study.

6.5 Summary

This work has presented a deep learning framework for identification of cattle based on muzzle point biometric feature. The deep learning approach has been used to learn the discriminatory feature of muzzle point images for better representation of muzzle point images with limited training dataset. The proposed deep learning approaches (CNN) [22] [106] , SDAE [187] [188] and DBN [81] [114]) provide the identification accuracy of 75.98%, 88.46%, and 95.99%, respectively.

The handcrafted texture features based representation algorithms is utilized for evaluations of experimental results. The Local Binary Pattern (LBP) and Circular-LBP (C-LBP) feature descriptor based techniques provide the rank-1 identification accuracy of $16.80 \pm 0.80\%$ to and $26.97 \pm 1.2\%$, respectively with four muzzle point images as gallery image per subjects (cattle). In case of appearance based feature extraction and representation approaches, such as principal components analysis is used to perform dimensionality reduction on the feature space.

The identification of cattle based on their muzzle point images is performed using Linear Discriminant Analysis (LDA) technique with One-Shot Similarity (OSS) technique. It is used to match a pair of samples and generate the match scores. The identification accuracies are shown in Table 6.6 and Table 6.7, respectively. It can be observed that Direct Kernel-LDA provides the $15.89 \pm 1.7\%$ to $29.97 \pm 1.13\%$ identification accuracy. The learnt feature descriptor techniques, such as VLAD+LDA+OSS and VLAD+LDA+SVM techniques yield $45.98 \pm 1.5\%$ to $59.64 \pm 1.12\%$ and $50.76 \pm 1.6\%$ to $67.98 \pm 1.17\%$

identification accuracy, respectively. Based on observation, we conclude that deep belief network deep learning approach provides better identification accuracy for recognition of individual cattle. Hence, it can be concluded that the DBN learning framework is the right choice for recognition purpose.

For further improvisation of deep learning based recognition framework, the proposed framework can be implemented in android platform that is available for smart or android devices for verification and recognition of false insurance claims in real time scenario.